°C °F
50 120
40 100
30 80
20 60
10 40
0 20
-10 0
-20 -20

GAMING THE METRICS

Misconduct and Manipulation in Academic Research

EDITED BY **Mario Biagioli** AND **Alexandra Lippman**

Gaming the Metrics

Infrastructures Series

Edited by Geoffrey C. Bowker and Paul N. Edwards

Paul N. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*

Lawrence M. Busch, *Standards: Recipes for Reality*

Lisa Gitelman, ed., *"Raw Data" Is an Oxymoron*

Finn Brunton, *Spam: A Shadow History of the Internet*

Nil Disco and Eda Kranakis, eds., *Cosmopolitan Commons: Sharing Resources and Risks across Borders*

Casper Bruun Jensen and Brit Ross Winthereik, *Monitoring Movements in Development Aid: Recursive Partnerships and Infrastructures*

James Leach and Lee Wilson, eds., *Subversion, Conversion, Development: Cross-Cultural Knowledge Exchange and the Politics of Design*

Olga Kuchinskaya, *The Politics of Invisibility: Public Knowledge about Radiation Health Effects after Chernobyl*

Ashley Carse, *Beyond the Big Ditch: Politics, Ecology, and Infrastructure at the Panama Canal*

Alexander Klose, translated by Charles Marcrum II, *The Container Principle: How a Box Changes the Way We Think*

Eric T. Meyer and Ralph Schroeder, *Knowledge Machines: Digital Transformations of the Sciences and Humanities*

Sebastián Ureta, *Assembling Policy: Transantiago, Human Devices, and the Dream of a World-Class Society*

Geoffrey C. Bowker, Stefan Timmermans, Adele E. Clarke, and Ellen Balka, eds., *Boundary Objects and Beyond: Working with Leigh Star*

Clifford Siskin, *System: The Shaping of Modern Knowledge*

Lawrence Busch, *Knowledge for Sale: The Neoliberal Takeover of Higher Education*

Bill Maurer and Lana Swartz, *Paid: Tales of Dongles, Checks, and Other Money Stuff*

Dietmar Offenhuber, *Waste Is Information: Infrastructure Legibility and Governance*

Katayoun Shafiee, *Machineries of Oil: An Infrastructural History of BP in Iran*

Megan Finn, *Documenting Aftermath: Information Infrastructures in the Wake of Disasters*

Laura Watts, *Energy at the End of the World: An Orkney Islands Saga*

Ann M. Pendleton-Jullian and John Seely Brown, *Design Unbound: Designing for Emergence in a White Water World,* Volume 1: *Designing for Emergence*

Ann M. Pendleton-Jullian and John Seely Brown, *Design Unbound: Designing for Emergence in a White Water World,* Volume 2: *Ecologies of Change*

Jordan Frith, *A Billion Little Pieces: RFID and Infrastructures of Identification*

Morgan G. Ames, *The Charisma Machine: The Life, Death, and Legacy of One Laptop Per Child*

Mario Biagioli and Alexandra Lippman, *Gaming the Metrics: Misconduct and Manipulation in Academic Research*

Malcolm McCullough, *Downtime on the Microgrid: Architecture, Electricity, and Smart City Islands*

Emmanuel Didier, translated by Priya Vari Sen, *America by the Numbers: Quantification, Democracy, and the Birth of National Statistics*

Ryan Ellis, *Letters, Power Lines, and Other Dangerous Things: The Politics of Infrastructure Security*

Gaming the Metrics

Misconduct and Manipulation in Academic Research

Edited by Mario Biagioli and Alexandra Lippman

The MIT Press
Cambridge, Massachusetts
London, England

© 2020 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC BY-NC-ND license.

Subject to such license, all rights are reserved.

This title is freely available as an open access edition thanks to the TOME initiative and the generous support of the University of California, Davis. Learn more at openmonographs.org.

This book was set in Sabon by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Biagioli, Mario, 1946- editor. | Lippman, Alexandra, editor.
Title: Gaming the metrics : misconduct and manipulation in academic research / edited by Mario Biagioli and Alexandra Lippman.
Description: Cambridge, MA : MIT Press, [2020] | Series: Infrastructures | Includes bibliographical references and index.
Identifiers: LCCN 2019010150 | ISBN 9780262537933 (pbk. : alk. paper)
Subjects: LCSH: Scholarly publishing—Corrupt practices. | Learning and scholarship—Corrupt practices. | Research—Corrupt practices. | Communication in learning and scholarship—Moral and ethical aspects.
Classification: LCC Z286.S37 G36 2020 | DDC 070.5—dc23
LC record available at https://lccn.loc.gov/2019010150

10 9 8 7 6 5 4 3 2 1

Contents

Introduction: Metrics and the New Ecologies of Academic Misconduct 1
Mario Biagioli and Alexandra Lippman

I Beyond and Before Metrics 25

1 Gaming Metrics Before the Game: Citation and the Bureaucratic Virtuoso 31
Alex Csiszar

2 The Transformation of the Scientific Paper: From Knowledge to Accounting Unit 43
Yves Gingras

3 Playing and Being Played by the Research Impact Game 57
Michael Power

4 The Mismeasurement of Quality and Impact 67
Paul Wouters

5 Taking Goodhart's Law Meta: Gaming, Meta-Gaming, and Hacking Academic Performance Metrics 77
James Griesemer

II Collaborative Manipulations 89

6 Global University Rankings: Impacts and Applications 93
Barbara M. Kehm

7 Predatory Publishing and the Imperative of International Productivity: Feeding Off and Feeding Up the Dominant 101
Sarah de Rijcke and Tereza Stöckelová

8 **Pressures to Publish: What Effects Do We See?** 111
Daniele Fanelli

9 **Ghost-Managing and Gaming Pharmaceutical Knowledge** 123
Sergio Sismondo

III **Interventions: Notes from the Field** 135

10 **Retraction Watch: What We've Learned and How Metrics Play a Role** 141
Ivan Oransky

11 **PubPeer: Scientific Assessment Without Metrics** 149
Boris Barbour and Brandon M. Stell

12 **The Voinnet Affair: Testing the Norms of Scientific Image Management** 157
Catherine Guaspare and Emmanuel Didier

13 **Crossing the Line: Pseudonyms and Snark in Post-Publication Peer Review** 169
Paul S. Brookes

14 **Ike Antkare, His Publications, and Those of His Disciples** 177
Ike Antkare

15 **Fake Scientists on Editorial Boards Can Significantly Enhance the Visibility of Junk Journals** 201
Burkhard Morgenstern

16 **Altmetrics Gaming: Beast Within or Without?** 213
Jennifer Lin

17 **Why We Could Stop Worrying About Gaming Metrics If We Stopped Using Journal Articles for Publishing Scientific Research** 229
Elizabeth Wager

IV **Mimicry for Parody or Profit** 237

18 **Making People and Influencing Friends: Citation Networks and the Appearance of Significance** 243
Finn Brunton

19 **Crack Open the Make Believe: Counterfeit, Publication Ethics, and the Global South** 251
Marie-Andrée Jacob

20 Fake Archives: The Search for Openness in Scholarly Communication Platforms 261
Alessandro Delfanti

21 Humor, Hoaxes, and Software in the Search for Academic Misconduct 271
Alexandra Lippman

Acknowledgments 283
Contributors 285
Index 287

Introduction: Metrics and the New Ecologies of Academic Misconduct

Mario Biagioli and Alexandra Lippman

Traditionally, the assumption has been that academic misconduct emerges primarily in response to "publish or perish" pressures. Robert Slutsky, a UC San Diego cardiologist famously caught in 1986 reporting imaginary experiments, was, at one point, putting out one article every ten days (Lock and Wells, 2001). "Publish or perish," however, is no longer the sole incentive for misconduct. New practices are emerging that are not limited to the production of fraudulent publications but are aimed instead at enhancing, often in unethical or fraudulent ways, the evaluation of their importance or "impact" (Biagioli, 2016). "Publish or perish" is merging with "impact or perish."[1]

This is related to but different from the predictable gaming of academic performance indicators one would expect from Goodhart's law: as soon as an indicator becomes a target, gaming ensues, which forecloses its ability to function as a good indicator.[2] That may take the form, for instance, of massaging the definition of what counts as a "successful student" in metrics about schools' performance, or of what counts as a "peer-reviewed" paper in faculty evaluation protocols. It could also involve aligning one's practices to metrics-relevant parameters, like capping classes' enrollment to nineteen students to have them fit the *US News and World Report*'s definition of "small class," which is rewarded in its ranking of universities. But we now find authors and editors who move beyond this kind of gaming to create (rather than tweak) metric-enhancing evidence, such as citations to one's work or to the work published in a given journal so as to boost its impact factor. We argue that the growing reliance on institutional metrics of evaluation does not just provide incentives for these kinds of manipulations, but also creates their conditions of possibility. They would not have come into being were it not for the new metrics-based "audit culture" of academia (Power, 1997; Strathern, 2000; Burrows, 2012).

Beyond Truth and Falsehood: Innovation in Manipulation

As shown by the US federal definition, misconduct is construed in epistemic terms: fabrication, falsification, and plagiarism.[3] Accordingly, misconduct is equated to producing false statements like making up data, fudging data, and faking authorship—false statements *within* a publication. Traditional fraud and misconduct continue to exist, and these definitions may be suitable to describe them. What they fail to grasp, however, are the new forms of manipulation that do not affect the epistemic status of a publication but take place *around and outside* the claims themselves like, for example, submitting fake peer reviews (often to publish in a higher impact factor journal than the article would have probably deserved), hacking journal databases (to manipulate the acceptance of one's article or to insert one's name in the authors' byline of an article already in press), setting up citation rings among authors (to maximize their personal citation counts) or among editors (to maximize their journals' impact factors), and so on. These may be called "post-production" manipulations in the sense that they concern the publication process and the impact of the claims, rather than a manipulation of the content of the publication. And while post-production manipulation may overlap with traditional misconduct, it does not need to. One can have a legitimate paper published in a journal with a good impact factor thanks to fake peer reviews.

We do not propose to simply expand old definitions of academic fraud to make room for post-production manipulations. Nor do we suggest that they are a lesser evil than traditional misconduct, or that they should be labelled as "questionable practices" rather than misconduct because they do not necessarily affect the core of a publication—its evidence and claims. The line between misconduct and questionable practice is notoriously hard to draw (Steneck, 2004; Biagioli, Kenney, Martin, and Walsh, 2019), and, more importantly, we cannot be positive that the different definitions of fraud and misconduct adopted by different countries, agencies, and academic institutions are accurate or fully commensurable with each other. This is not meant as a criticism but as an acknowledgment that the thinking, definitions, and policies about misconduct have been and continue to be the work in progress of hundreds of concerned practitioners in universities, governmental institutions, funding agencies, and journals (Jacob, 2014).

Our goal is neither to question nor to uphold existing views about traditional misconduct, but to call attention to and provide a first analysis of a recent dramatic development: the emergence of a range of new ways of manipulating the publication process—manipulations that are qualitatively different from traditional academic misconduct. Many of the practices described in this book are obviously unethical and arguably illegal. For instance, independently of the nature of the specific manipulation involved in a given case, the goal of most of these practices is to produce an artificially enhanced curriculum vitae, which gives its holder a leg up against fair-playing competitors in gaining positions and funding that they might not have otherwise obtained. This would seem to match the legal definition of fraud, without having to mobilize more specialized definitions of academic or scientific misconduct. Our primary interest as scholars, however, is to understand the features of these new forms of manipulation, why they are emerging now, what motivates or incentivizes them, and what are the new forensic techniques and actors that are being mobilized to detect them. Ultimately, these questions are key to determining whether we are confronting new instances of old misconduct—old wine in new bottles—or something altogether different. In old Kuhnian parlance, the fact that several of these practices seem irreducible to the current misconduct taxa (fabrication, falsification, and plagiarism) does not tell us whether they are merely apparent anomalies that can be eventually massaged into our current "misconduct paradigm," or whether, instead, we might end up having to rethink it altogether. We believe that the latter outcome is the more probable one, and hope that our book will contribute to that rethinking by providing a comprehensive map of the problem and its contours. As our many chapters show, the emergence of this new species of misconduct (or whatever more appropriate term we might develop down the line) should be seen simultaneously as a problem and a symptom of a more general shift in the academic publication system, down to the very meaning of publication, and thus of misconduct.

To put it somewhat crudely, what we discuss here are less "epistemic crimes" than "bureaucratic crimes"—practices involving the production of publications that manipulate the publishing system itself or, more specifically, what that system has recently evolved into. What is clear is that these manipulations amount to post-production activities and that, despite their many different forms, they are framed by metrics of evaluation variously based on citations and impact rather than by concerns with plain productivity (as it was in the "publish or perish" age). Conflating epistemic and bureaucratic manipulations would risk foreclosing an understanding of the conditions behind the emergence of these new practices and what that may tell us about what "publication" is becoming in the age of metrics in an increasingly global context.

Redefining Publication and Evaluation

In the seventeenth century, “publication” meant making things public by a variety of means, ranging from lectures (to students or fellow academicians) to letters, personal conversations with reputable people, and printed publications. That changed in the nineteenth century when, at least in science, the definition of publication was narrowed down to printed articles in academic journals (Csiszar, 2018). But if the definition of acceptable modalities and technologies of publication changed in time, the assumption remained that the evaluation of the claims made public was to be done by those who listened to or read them. It was a judgment made by humans, which could be contested by challenging either the protocols employed in the judgment, or the qualifications of the people making them. Today, instead, publication is no longer necessarily evaluated through reading by people but, in some contexts, through markers connected to the publication (though external to it), such as the impact factor of the journal where it is published or the number (or, in some cases, the weight) of the citations it receives (West et al., 2010; Biagioli, 2018) or article metrics on platforms like Academia.edu and ResearchGate (Lippman and Kelty 2019).

The meaning of “publication” has substantially changed, not just because its evaluation has almost ceased to require human agency, but also due to the fact that publication is no longer limited to the process of making claims public. Publication used to be separate from evaluation (which was clearly thought of and practiced as a post-publication activity), but the two may now be folded together. When it relies on the journal’s impact factor, evaluation no longer follows publication but takes place *together with* the act of publication. It involves locating the venue of the publication and attaching that location’s index—the impact factor—to the publication. A publication is born evaluated, so to speak.

The meaning of evaluation has changed as much as that of publication. It is not just that, as we often hear, nobody reads but people only count. Something more radical has happened: the traditional locus of evaluation—the publication’s claims—has become technically irrelevant to metrics regimes based on impact factors. It is not that people ought to read but have lazily stopped doing that. People still read for research and educational purposes, but reading is no longer a necessary component of *institutional* forms of evaluation because some of those metrics are independent of the epistemic dimensions of that specific publication—its claims—but rely, instead, on metadata and similar markers that can be picked out and processed by nonhumans.

From Content to Metadata

These changes in what "publication" and "evaluation" mean are not immediately evident from the external appearance of an actual publication. Whether you look at it in print or online, a journal article still looks very much like it did in the 1980s. What has changed is the role of its metadata, which has gone from descriptive to evaluative. The title of the journal where the article is published no longer simply describes where that article became public (or where it can be found on a library's shelf), but, through the impact factor, it conveys a numerical estimation of its reception and effect. Conversely, an article published in a journal whose title appears on a list of online open-access "predatory journals" (discussed below) may be simply dismissed as having no value whatsoever by a committee reviewing a researcher for hiring or promotion. It could be effectively treated as a nonpublication despite the fact that it looks exactly like an article published in a journal that looks exactly like a journal. It is the shift of the focus of evaluation from the claims (internal to the publication) to the circumstances of the publication (external to it) that anchors all the changes discussed above: the end of reading (replaced by the scraping of metadata); the switch from qualitative human judgment to quantitative calculated indexes; and the merging of publication and evaluation.

It is worth noticing that, taken as an institutional genre, these new modalities of evaluation are an expansion of the form of library cataloguing. Like cataloguing, they do not involve the reading of a publication's content but rather the processing of a publication's metadata. We might say that impact-oriented evaluation becomes part of an "expanded indexing," one that does not simply generate a call number but processes aggregate publication metadata to generate figures about the purported value of the publication as an input in a variety of institutional decisions—faculty hiring and promotion; whether the library should or should not subscribe to that journal (or to the catalogue of a given publisher); whether grants to defray open-access publication costs should be given to faculty who want to publish in those journals; and so on. If in the past the evaluation of a publication was almost exclusively undertaken to either assess the quality of its claims or the scholarly quality of its author, the new forms of evaluation based on indexes (rather than content) are aimed at informing a variety of institutional decisions, down to the national or even global ranking of the institutions themselves.

Are Journals Becoming Mints?

Also striking is what has become of the notion of impact. In common usage, impact refers to an effect, that is, to something that has happened already (like, say, the citations that an article has received since its publication). The increasingly coveted Journal Impact Factor (JIF), however, functions as an estimation of impact before it happens, as a device to give a valuation right now to a publication that can in fact accrue value (that is, impact) only in the future. This is different from saying that the value of things is bound to fluctuate in time. A house has value both when it is first built and then years after that, but, by definition, the impact of a publication is *zero* when the publication comes off the press. This means that a publication's impact cannot be measured by the impact factor because there is literally nothing to measure at the time of the publication. All the JIF can produce is purely a prediction of impact, and one that is not based on the features of that specific publication but on those of previous articles published in that journal over a certain period of time.

Like actuarial or death tables, the impact factor is based on evidence about the past. And in the same way that actuarial tables are used to estimate the likely length of one's future life to calculate today's insurance premium, the impact factor is deemed to provide an estimation of the amount of citations the article *will have received* in the future based on the fact that it was published in that journal, thus "pricing" the article (and thus its authors) *right now* rather than after it would have had actual impact. It is a rather crude tool to price futures. We are not speaking metaphorically: In China, universities hand substantial cash bonuses to their faculty for their publications, indexing them to the journals' impact factor. *Nature* and *Science* articles fetch over $30,000 a piece.[4] Similar schemes can be found on the other side of the Equator:

> Melbourne Business School pays $A15,000 cash for every paper published in the Top 40 list compiled by the *Financial Times*. The scheme at Queensland University Business School is more complicated. Payments, made to the departmental accounts of authors, are approximately: Tier 1 journal—$A12,000; Tier 2 journal—$A7500; Tier 3 journal—$A5500; Tier 4 journal—$A2000; Professional journals—$A1000. (Macdonald and Kam, 2007)

There is nothing wrong in using reasonable estimates about future states of affairs, except that the impact factor is not used as an estimate, but has been reified into a positive measure of impact and, more broadly, of value. Furthermore, while the impact factor refers specifically to the journal (not the articles), the JIF has come to signify the impact or value of each article published by that journal. It functions, literally, like money. It is the "face value" impressed on a coin or banknote, which determines its value no matter what the coin is made of (copper, silver, gold, steel), or what the exchange value that coin would have as metal. The "content" of the coin is just the medium for the stamp, which needs to be impressed on some material. What carries value is not the inside but the outside—the number inscribed on the surface. Whether the article is a piece of gold or lead (or worse), the JIF has come to determine the face value of that article. (We could probably think of journals as mints printing or coining money with a face value equivalent to their impact factor.[5]) I do not even need to know what the JIF is, how it is calculated, or how reliable it is because my institution—and probably all institutions I will ever work for—will honor that face value. An article in *Science* or *Nature* is literally money in the bank, independently from the actual impact it will ever have.[6]

Attaching an impact factor to an article at the time of publication (and thus before real impact has accrued) shows that the JIF has literally nothing to do with the evaluation of that specific article, but simply prices it in a currency that allows for *exchange*. As an author, I can "trade" articles with certain impact factors into a job, and the institution that employs me can then "trade" those publications (together with hundreds or thousand more by other faculty of the same university) into a better national or global ranking, which may be subsequently traded into more students, donors, contracts, and so on. (Conversely, these days in China, one can purchase authorship in a prewritten and preaccepted article to be published in an English language journal, at prices that vary according to the impact factor of the specific journal.[7]) The impact factor (which we are using as an exemplar of metrics) is thus literally neither about the evaluation of a *specific* article nor about making evaluation fair and transparent by removing it from the arbitrariness of qualitative judgments: It is about *creating the conditions of possibility for a market.*

No matter how accurate available actuarial tables may be, one can hardly develop a life insurance industry without them. And even if they are statistically good, they are still very unlikely to accurately predict the exact date of death of Ted the baker around the corner (in the same way that even the best of impact factors is not going to correctly predict the impact of any specific article). But that's not the point. It does not matter that, at best, the impact factor can only capture some features of a *population* of articles published in a given journal. The role of the JIF is spreading (despite some spirited opposition)[8] because it produces prices and a currency in which those priced goods can be exchanged and circulated between authors, universities, libraries, and publishers. And, crucially, these prices can be determined at the time of publication, without having to wait to count the citations it will accrue in time. The journal impact factor shaves off the several years it would have taken for that article to grow its value, thus enabling more scholars—especially junior ones—to enter the market with something that looks like "hard value" rather than what they mostly have, that is, possible value in the future.

Emergent Objects, Emergent Manipulations

This shows, yet again, that post-production manipulations like those aimed at impact factors are qualitatively distinct from traditional misconduct, making it difficult to define what kind of misconduct they are, and whether misconduct is indeed the right term. They are not about manipulating knowledge claims but their institutional valuation. This means that, unlike traditional misconduct that has been boiled down to three practices—falsification, fabrication, and plagiarism—post-production manipulations can take up as many forms as there are metrics techniques and markets, which are constantly changing. The dramatic innovations in post-production manipulations that have emerged over the last few years support this view, while also suggesting that post-production manipulation may not be an object stable enough to be definable. New metrics and indicators are being constantly introduced or modified often in response to the perceived need to adjust or improve their accuracy and fairness. They carry the seed of their never-ending tuning and hacking, as each new metric or new tuning of an old one can be subsequently manipulated in different ways. Also, new "markets" and uses keep developing for existing metrics, which means that new categories of actors can get in the game, manipulating the metrics in different directions in response to specific goals. Metrics of student performance evaluation may become, for instance, tools to evaluate the performance of teachers, or of an entire school district. Closer to home, we know (as Csiszar's chapter shows) that scientometrics emerged as a tool for mapping scientists' works and networks, but was then turned into a tool for evaluating them. Similarly, the JIF was meant to evaluate journals, but has become the premier tool to assess the value of articles. The techniques and indicators may remain the same, but the changing significance of the outcome can be enough to create a market for new manipulations.

In sum, we do not yet have a new concept that can capture all the various manifestations of post-production misconduct, and it is not likely that we will develop one anytime soon. All we know is that all these forms share the same *telos*: the manipulation of the metrics of academic evaluation. The means take many different forms, but the ends stay the same.

Ways of Gaming

Unlike traditional scientific fraud, post-production manipulation is no longer the purview of individuals but rather involves groups, networks, or entire institutions. Journal editors conspire to increase their publication's JIF through co-citation agreements among journals, authors organize themselves into fake peer-review rings, and editorial service providers not only help scientists write their articles in good English, but also, for an extra fee, help to line up friendly peer reviews. Up the institutional ladder, universities select or massage their data to score well on global university rankings (whose importance is growing with the increasing global scale of the higher education market).

Post-production manipulation has not only become a more collaborative effort, but it has also moved beyond the sites of traditional fraud like universities, corporate laboratories, and federal research institutions. It has spread from places where research originates to places where it goes to, like journals (especially those which librarian Jeffrey Beall and others termed "predatory journals"[9]) and "fake" conferences in vacation destinations that may use impressive-looking but possibly fake advisory committees while promising to publish the papers' abstracts (likely accepted without review) in journals that would probably fall in the "predatory" category (Brooks, 2009). You went on vacation (possibly paid for by your grant or research funds) and came back to find your vitae enriched by one additional conference talk and a publication.

Traditionally, journals have been cast in the role of gatekeepers, credited with the ability to sort good science from bad through peer review. Today, however, we see so-called "predatory journals" actively contributing to the post-production manipulation trend. While criticized mostly for their virtual freedom from the constraints of peer review, for their pay-to-play business model, and for their tendency to have fictional editorial boards (Morgenstern, this volume, chapter 15), we find the emergence of this breed of journals particularly interesting as a window on the logic of the new metrics-based regime of science publications. While light-years away from high impact journals like *Science*, *Nature*, or *Cell*, these "predatory journals" may be simply the other side, or perhaps the bottom, of the *same* metrics economy. Their impact factor is often insignificant (or possibly made up), and yet they seem to provide a crucial service—possibly even a lifeline—to authors who are struggling to meet the quantitative publication benchmarks set by their institutions, or their strict deadlines.

These authors' institutions may be neither high ranking nor particularly ambitious, but they can hardly ignore the ranking game, which has become global and played at all levels of the "excellence" spectrum. As a result, these universities may still demand their faculty to publish a certain number of publications in nonlocal journals, that is, "international" English-language venues (de Rijcke and Stöckelová, this volume, chapter 7). The content of the article may not be crucial if the author can at least appear to be productive and able to publish in English—a fact that his/her department chair and dean can turn into a figure they can use as they pitch the steps—however modest they may be—that they are taking toward leading their institutions on the long path to excellence.

What these journals produce, therefore, are not publications but *publication effects* or publication *tokens*. More objects than texts, these publications (if publication is indeed the proper term here) are not meant to be read, but are rather generated for the sole purpose of allowing the authors to add an entry to their curriculum vitae, or for their deans to tally and include them in their annual reports to the higher administration or to prospective donors.[10] These publications and journals may be virtually impact free, but they are not outside of the metrics-based economy of impact. They are simply at the bottom of that economy and, no matter the scorn they receive, no economy can function without a bottom.

The rise of "impact or perish" has also been accompanied by a marked increase in journal self-citation. In some cases, editors pursue this by asking prospective authors (especially junior ones) to cite other articles from the journal they have submitted their articles to (Wilhite and Fong, 2012). Clarivate Analytics (formerly Thomson Reuters) tracks self-citation and bans journals whose self-citation is deemed excessive. For instance, after the *Journal of Biomolecular Structural Dynamics*'s impact factor spiked from 1.1 to 5.0 in just one year between 2009 and 2010 (Van Noorden, 2012), Thomson Reuters asked the journal to explain its success. *JBSD*'s editor-in-chief attributed the journal's sharp rise in impact factor to their publication of a controversial paper, which generated many responses, and to a new policy encouraging authors to explain their work's connection to other articles previously published in the journal (Van Noorde, 2012). Unconvinced, Thomson Reuters excluded the journal from its *Journal Citation Reports* (the annually updated list of the impact factors of the journals tracked in the Web of Science database), effectively denying them a ranking. But while such bans are becoming increasing common (rising from nine in 2007 to thirty-eight in 2013), some believe they should be imposed more frequently given that even Thomson Reuters's "own statistics indicate that 140 journals have had self-citations making up more than seventy percent of total citations in the past two years" (Van Noorden, 2012).

If self-citation is easy to track, citation cartels between journals are significantly more difficult "to detect since they represent coordinated efforts among several journals to collectively self-cite."[11] Perhaps emboldened by this gap in the available technologies of detection, in 2009, the editors of eight Brazilian science journals decided to boost their impact by agreeing to publish "articles containing hundreds of references to papers in each others' journals" (Van Noorden, 2013). The scheme worked well for a few years, until 2013, when Thomson Reuters developed an algorithm to detect citation rings by spotting "concentrated bursts of citations from one journal to another."[12] This led them to suspend fourteen journals from the *Journal Citation Reports*, including four members of the Brazilian citation ring.[13] One of the editors explained that:

> The citation ring grew out of frustration with his country's fixation on impact factor. In Brazil, an agency in the education ministry, called CAPES, evaluates graduate programs in part by the impact factors of the journals in which students publish research. As emerging Brazilian journals are in the lowest ranks, few graduates want to publish in them. This vicious cycle, in his view, prevents local journals from improving. (Van Noorden, 2013)

Another member of the ring—the editor of the *Jornal Brasileiro de Pneumologia*—emphasized a link between the impact factor and the global politics of publication: the scheme to boost the journals' impact factor was not only self-serving but "also to show off articles in Brazilian journals, attracting better contributions and raising quality all round" (Van Noorden, 2013). Whether self-serving or patriotic, the scheme was surely incentivized and made possible by the increasing global "hegemony" of the impact factor (Barbour and Stell, this volume, chapter 11).

Some scholars who wish to bypass the constraint of having to either write papers or plagiarize them from other scholars opt to use SCIgen—an article-generating software. This is rather ironic, or worse, given that the developers of SCIgen—three graduate students at MIT—created this software as a way to expose, rather than contribute to, unethical evaluation and publication practices. Tired with the invitations to spurious conferences and journals that clogged their inboxes, they created SCIgen to generate nonsensical (but legitimate-looking) submissions, which they then fed to "fake" conferences like the capaciously titled "World Multi-conference on Systemics, Cybernetics, and Informatics." Their goal was to expose the fact that such conferences accepted any paper that came their way, without review, provided the presenter was willing to pay the registration fee. But, in a move that ran diametrically counter to their critical and humorous intent, the software was quickly co-opted by scientists who started to use it for real, effortlessly cranking out papers that they then humorlessly submitted to conferences—conferences that, as the SCIgen team had suspected, did accept and subsequently publish them (Antkare, this volume, chapter 14). (This is yet another example of how these new forms of manipulation seem to expand by repurposing tools to enable new opportunities, and so on.)

SCIgen-generated texts traveled far and wide. After creating software to detect such articles, Cyril Labbé (Lippman, this volume, chapter 21) identified and catalogued scores of computer-generated papers published in the proceedings of respectable conferences, not just the spam-like ones that the original SCIgen pranksters wanted to make fun of. For instance, Springer and the Institute of Electrical and Electronic Engineers (IEEE) had to retract more than 120 papers they had previously accepted and published (Van Noorden, 2014; Bohannon, 2015). Following the embarrassment, Springer enlisted Labbé and his laboratory to develop SciDetect, an open-source software to "ensure that unfair methods and quick cheats do not go unnoticed."[14] (Parenthetically, two years ago, a team at the University of Trieste introduced software to produce fake peer reviews. Like their SCIgen brethren, they mean it as a prankish tool to expose the problems of peer review, but only time will tell if it will also be used for real.[15])

Peer reviews too can be manipulated in various ways. The website Retraction Watch (Oransky, this volume, chapter 10) has reported many cases—more than six hundred papers so far[16]—of rigged peer review. These are cases in which authors submitted email addresses of suggested reviewers that were in fact registered to the authors themselves. When the journals took up the suggested reviewers (which happened more often than one would expect), the authors received an email inviting them to review their own papers, which they typically found promising and publishable with a few revisions. In more sophisticated schemes, authors added citation rings to these rigged peer reviews (Ferguson et al., 2014). Finally, the Committee on Publication Ethics (COPE) has found that some organizations sell services "ranging from authorship of prewritten manuscripts to providing fabricated contact details for peer reviewers during the submission process and then supplying reviews from these fabricated addresses."[17]

The rising importance of the impact factor has also created a market for fake ones, which are particularly attractive to low-quality journals trying to look better than they are. Beall warned that, "in this competitive market, publishers want to stand out from the crowd and attract the author fees. One way to effectively earn these fees is to boast high journal rankings"[18]—rankings that one can buy pretty much off the shelf. Many of the "tailored" impact factor providers mimic legitimate scientometrics services through similar-sounding names and website domains. For instance, the shamelessly counterfeit "Thomson Reuters Institute for Scientific Information" ran the website www.isi-thomsonreuters.com (as distinct from the original www.thomsonreuters.com) and claimed to be the actual ISI (the original scientometrics company). Other equally creative citation companies supplying impact factors to so-called predatory journals include Universal Impact Factor (UIF), Global Impact Factor (GIF), and Citefactor (Jalalian, 2015). By providing impact factors for established journals such as *PLoS* and *Nature* along with less-reputable journals, fake impact factor companies contribute to the appearance of legitimacy within the ecosystem of post-production manipulations.

Finally, while universities use metrics to evaluate their faculty, they are subjected, in different contexts, to some of those same metrics (Espeland and Sauder, 2016). Several global university rankings, for instance, consider the faculty's citation counts. In recent years, universities have sought to raise their rankings by targeting the very indexes by which they are ranked, and hiring consultants to identify exactly how to do that. Northeastern University provides one of the most successful examples of how to target the *U.S. News and World Report* rankings. Already in 1996, former Northeastern University President Richard Freeland observed "how schools ranked highly received increased visibility and prestige, stronger applicants, more alumni giving, and, most important, greater revenue potential. A low rank left a university scrambling for money. This single list…had the power to make or break a school" (Kutner, 2014). Following this insight, Freeland identified precisely what Northeastern would need to target—class size, graduation rate, admission statistics, and so on—to rise in the rankings and break into the top 100. The effort paid off, and then some. From its score of 162 in 1996, Northeastern rose to 98 in 2006, and to 44 in 2018. This is by no means an isolated case. Other universities' tactics for swaying rankings have included hiring top-cited faculty as well-paid, part-time, affiliated, nonresident faculty,[19] inflating students' SAT scores, high school GPAs, and graduation rates, and refunding poorly performing freshman students (or, in Yale's case, first-year law school students) their tuition after their first semester if they are willing to drop out.

Fake Is No Longer What It Used to Be

We find it intriguing that the results of post-production manipulations developed to meet or exceed the performance benchmarks established by institutions, funding agencies, or global university rankings cannot be easily categorized as "fake" in the same way that traditional misconduct could be said to involve fake evidence or fake authorship. For lack of a better term, we may use "spammish" to point to a tension inherent in some post-production manipulations that is not captured by terms such as "fake," "bogus," or "predatory."

No doubt, there are differences between a traditional conference and a for-profit event organized largely for the purpose of having scientists pay to deliver papers and have them printed in some obscure journal (with possible ties to the conference organizers) (Grant, 2016; Straumsheim, 2016). Still, it would be inaccurate to say that the latter is simply a "fake" version of the former. Similarly, obscure journals whose editors clog our inboxes with offers to publish next week what we submit by this Friday are definitely suspect, but not merely "fake." They surely engage in misrepresentations (as when they boast stringent peer-review standards), but they mostly withhold information from their prospective authors, creating ambiguities that play in their favor. They are still academic journals, some of them listed in the standard indexes (though often in fewer than the ones they boast). And they do indeed publish articles, some of which get cited. Occasionally, some of these journals are bought up by prime-time publishers, suggesting that they may be perceived as "emergent" rather than simply "fake." (Unless of course you think that the big publishers buy up these journals simply because they are profitable, without worrying too much about their publishing ethics).

Similarly, while some of the scholars whose names grace their editorial boards may not be aware of being listed there (see Morgenstern, this volume, chapter 15), that does not mean that "fake" fully describes those boards. In some cases, the editorial boards are indeed made up, but, in others, advisors agree to have their names listed, perhaps because they do not understand the nature of those journals, or because they want to support affordable open-access publishing.[20] In other instances, however, they may decide not to ask too many questions and simply accept the invitation so as to add another line to their vitae.

Again, while there is a certain fakeness about these journals, to say that they are plainly “fake” or “predatory” misses the complexity of the ways in which they are both fake and not fake, and the fact that such ambiguities are central to their role and business model. While it is difficult to find kind things to say about these journals, their relentless vilification as illegitimate, fraudulent, and rapacious looks like othering—an index of how the publishing professionals of the Global North use these journals as a foil to project a desirable identity and image of themselves. (As some of the chapters argue, these “predatory” journals may be in fact a blessing in disguise for the more established publishers by providing a “bad” benchmark against with they can strut their quality.) Furthermore, while the generic and random nature of the invitations we receive to publish in these journals, the wildly capacious assemblages of disciplines that are conjured in their titles, or the impressively fast publication time they boast seem as far fetched as the large sums of money that spam emails claim to be waiting for us in some remote bank account, these journals cannot be said to be truly “spam” or “predatory” either (Brunton, this volume, chapter 18). It would take a truly naïve scientist to believe that these publication venues belong in the same category and have the same credibility of the peer-reviewed journals in which they would rather publish.[21] Equally naïve would be the belief that their submissions could be properly reviewed and published in a matter of days, or that the emails one receives from these editors (populated by strange typos and dubious academic links) are actually coming from those people and from those addresses. Given that academics are a reasonably intelligent bunch, those who choose to publish in these venues are not likely to be deceived into doing so. An attempt to deceive is surely involved here, but one too transparent to justify saying that those who accept those invitations are “preyed upon,” cheated out of the money they send to these journals for publication costs or for gold open-access fees (which, in any case, are a fraction of the going rates of more established journals).

“Spammish” may be a better term to capture the ambiguous nature of these practices that, while appearing spam-like, are to some extent collusive. It may be to the authors’ advantage to treat publications in these suspicious journals or attendance to these suspicious conferences as perfectly legitimate and worthy of inclusion in one’s vitae, only to say that they got tricked into publishing in these journals or going to those conferences in case their university questions their choices. Neither "real" nor "fake," these practices are effective precisely because they can be played and presented differently, depending on the circumstances (Jacob, this volume, chapter 19).

New Evidence, New Watchdogs

It should not come as a surprise that modes of misconduct detection have changed with the emergence of new metrics-based post-production manipulations, and the changing notions of fakeness that go with them. While some of these manipulations may still be detected through peer review, most of them are beyond the reach of traditional referees. In some cases, peer review is unable to function as a gatekeeper precisely because it becomes itself the target of those manipulations. (For instance, peer review and citation rings function by rigging peer review, that is, by replacing legitimate reviews with counterfeit ones.) Some of these new manipulations, therefore, can be detected not through careful reading of a manuscript, but only through extensive analysis of journal databases—of the wording of reviews, review turnaround times, citation patterns, and the mutual relationships between authors and reviewers across different publications.[22]

This requires both a different kind of expertise as well as access to different levels of evidence and data. Much of this evidence, in fact, can be mined only by teams of investigators, hired by the publishers, carefully poring over information held in proprietary journal databases. And as shown by the terseness and brevity of most retraction notices, editors and publishers are often reticent to expose how their editorial processes have been gamed, as that exposes weaknesses in their systems and services (Biagioli, 2016). Less than ten years ago, a now-prominent misconduct researcher contacting a journal about an uninformative retraction notice was told that, "it's none of your damn business"—a kind of answer that some editors still relay today (when they respond, which they do less than half of the time).[23] We are seeing, in sum, both an increase in the amount of forensic traces of misconduct as well as the decrease of the readability of those traces, which are now often beyond the reach of traditional peer review. This "privatization" of forensic evidence and the proliferation of its forms have been paralleled, however, by an opposite trend: the crowdsourcing of the discussion and analysis of evidence of potential misconduct.

The emergence and pervasiveness of new forms of misconduct exceeds the reach, resources, and conceptual framework of traditional governmental watchdog organizations like the Office of Research Integrity in the United States—agencies that are already undergoing some identity crisis (Kaiser, 2016). But these institutional bodies are no longer the sole players. Noninstitutionally affiliated watchdog organizations have emerged, like Retraction Watch, PubPeer, blogs such as Scholarly Open Access, and other sites like the now-defunct Science Fraud (Pain, 2014; Aschwandan, 2015; Blatt, 2015). This new generation of watchdogs is successfully making up for their lack of resources by mobilizing hundreds of scientists—some named, but mostly anonymous—who are willing to read texts, evaluate images, run through statistical analyses of a publication's data, and share their findings and views on websites, blogs, wikis, and social media. As Eric Raymond famously said about open-source software collaborative practices, "given enough eyeballs, all bugs are shallow" (Raymond, 1999). And though they lack legal authority, these new watchdogs can be very effective through their ability to maximize the visibility of these issues, which may force the authorities to intervene (Guaspare and Didier, this volume, chapter 12). Interestingly, all these efforts have been moving in the opposite direction from the forms of evaluation characterized by the adoption of the impact factor and other metrics. Rather than looking at metadata, these watchdogs engage in careful reading of the content of the publication—data, images, text—and sometime attempt to reproduce the claims. Their modus operandi is that of traditional peer review but, through the adoption of a crowdsourcing model, it operates on a scale and is able to draw expertise from a population that is an order of magnitude larger than that of traditional peer review as practiced by journals. This change of scale has a direct impact on the granularity of the review, but it also profoundly shifts the "governance" of misconduct research from governments and institutions to the practitioners themselves.

And as these developments mark a transformation from top-down to bottom-up knowledge production, they are often accompanied by an affect that is rather unusual in academia: humor. There is a clear migration (Lippman, this volume, chapter 21) away from the high seriousness and humorlessness of the discourse of university committees and governmental agencies such as the ORI, and toward the carnivalesque attitude of some of the new watchdogs. Wearing the masks of anonymity and pseudonimity, and cracking jokes (some better than others), they blur the line between "policemen" and "pranksters."

Notes

1. In some cases, we can empirically see the shift from one to the other. In places where the number of publications is the top target, people publish much but not in particularly high-impact journals. Instead, where institutions tie rewards to publications in top-tier journals rather than to sheer quantity, authors adjust by publishing less but in higher impact journals.

2. "Goodhart's law" is named after Charles Goodhart who formulated it in a paper from 1975. For an example of an analysis of "Goodhart's law," see Marilyn Strathern, "'Improving Ratings': Audit in the British University System," *European Review* 5(1997):305–321. The adage is also sometimes referred to as "Campbell's law" from Donald T. Campbell's 1975 paper, "*Assessing the Impact of Planned Social Change*" or the "Lucas Critique," which economist Robert Lucas articulated in "*Econometric Policy Evaluation: A Critique*" (1976). The relationship between "Goodhart's law" and the "Lucas Critique" is discussed in Alec Chrystal and Paul Mizen, "Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy," in Paul Mizen, ed., *Central Banking, Monetary Theory and Practice*, Vol. I (Cheltenham: Edward Elgar Publishing, 2003), pp. 221–243.

3. The Office of Research Integrity, "Definition of Research Misconduct," Accessed September 1, 2016, http://ori.hhs.gov/definition-misconduct.

4. Phil Davis, "Paying for Impact: Does the Chinese Model Make Sense?," *Scholarly Kitchen*, April 7, 2011, Accessed August 1, 2016, https://scholarlykitchen.sspnet.org/2011/04/07/paying-for-impact-does-the-chinese-model-make-sense. Jufang Shao and Huiyun Shen, "The Outflow of Academic Papers from China: Why Is It Happening and Can It Be Stemmed?," *Learned Publishing* 24(2011):95–97.

5. This analogy brings up interesting questions about the difference between printing (money) and publishing knowledge, and how the JIF is effectively narrowing that difference, if not blurring it outright.

6. Encouraging scholars to submit to high-impact, international journals also harms the development of high-quality, higher impact Chinese journals according to Wang Shuhua and Paul Weldon, "Chinese Academic Journals: Quality, Issues, and Solutions," *Learned Publishing* 19(2006):97–105. See also Xu Jie and Matthias Wahls, "The Scholarly Publishing Industry in China: Overview and Opportunities," *Learned Publishing* 25 (2012):63–74.

7. Alison McCook, "7 Signs a Scientific Paper's Authorship Was Bought," *Retraction Watch*, October 10, 2016, Accessed October 27, 2016, http://retractionwatch.com/2016/10/24/seven-signs-a-paper-was-for-sale/. Mara Hvistendahl, "China's Publication Bazaar," *Science* 342(6162):1035.

8. See, for instance, Phil Davies, "On Moose and Medians (or Why We Are Stuck with the Impact Factor)," *Scholarly Kitchen*, April 12, 2016, Accessed August 15, 2018, https://scholarlykitchen.sspnet.org/2016/04/12/on-moose-and-medians-or-why-we-are-stuck-with-the-impact-factor.

9. For years, Beall has maintained an online blacklist (now removed) and a widely read and sometimes-critiqued blog, Scholarly Open Access. See for instance, Monica Berger and Jill Cirasella, "Beyond Beall's List: We Need a Better Understanding of Predatory Publishing Without Overstating Its Size and Danger," *The Impact Blog*, March 18, 2015, Accessed June 26, 2016, http://blogs.lse.ac.uk/impactofsocialsciences/2015/03/18/beyond-bealls-list-predatory-publishers.

10. Mario Biagioli, "Recycling Texts or Stealing Time? Plagiarism, Authorship, and Credit in Science," *International Journal of Cultural Property* 19(2012):463–464.

11. Phil Davis, "Citation Cartel Journals Denied 2011 Impact Factor," *Scholarly Kitchen*, June 29, 2012, Accessed June 15, 2016, https://scholarlykitchen.sspnet.org/2012/06/29/citation-cartel-journals-denied-2011-impact-factor. On recent developments toward detecting citation cartels, see Iztok Fister et al., "Toward the Discovery of Citation Cartels in Citation Networks," *Frontiers in Physics*, December 15, 2016, Accessed 8/1/2018, http://journal.frontiersin.org/article/10.3389/fphy.2016.00049/full.

12. Ibid.

13. *The Journal Citation Reports*. Accessed June 1, 2016, http://adminapps.webofknowledge.com/JCR/static_html/notices/notices.htm.

14. Renate Bayaz, "Springer and Université Joseph Fourier Release SciDetect to Discover Fake Scientific Papers," March 23, 2015, Accessed August 15, 2018, https://www.springer.com/gp/about-springer/media/press-releases/corporate/scidetect/54166.

15. Jack Grove, "Robot-Written Peer Reviews," *Inside Higher Ed*, September 22, 2016, Accessed September 23, 2016, https://www.insidehighered.com/news/2016/09/22/many-academics-are-fooled-robot-written-peer-reviews#.V-Qd_s9vpF0.gmail.

16. Allison McCook, "A Publisher Just Retracted Ten Papers Whose Peer Review Was 'Engineered—,'" *Retraction Watch*, July 12, 2018, Accessed August 14, 2018, https://retractionwatch.com/2018/07/12/publisher-has-known-of-problem-of-fake-reviews-for-years-so-how-did-10-papers-slip-its-notice/.

17. "COPE Statement on Inappropriate Manipulation of Peer Review Processes," December 19, 2014, Accessed June 20, 2016, http://publicationethics.org/news/cope-statement-inappropriate-manipulation-peer-review-processes. Mara Hvistendahl has also reported on "China's Publication Bazaar" in which agencies, scientists, and editors sell papers indexed by Thomson Reuters and Elsevier in "China's Publication Bazaar," *Science* 342(6162):1035–1039.

18. Jeffrey Beall, "Look Out for Bogus Impact Factor Companies," *Scholarly Open Access*, Accessed July 1, 2016, https://scholarlyoa.com/2013/08/06/bogus-impact-factor-companies. See also Mehrdad Jalalian, "The Story of Fake Impact Factor Companies and How We Detected Them," *Electronic Physician* 7(2):1069–1072.

19. Lior Pachter wrote about how unknown King Abdulaziz University hired top-cited mathematicians as "distinguished adjunct professors" and rose to number 7 in USNWR's global university ranking by subject. Accessed July 1, 2016, https://liorpachter.wordpress.com/2014/10/31/to-some-a-citation-is-worth-3-per-year.

20. Jeffrey Beall, "Bogus Polish Journal Has Completely Fake Editorial Board," *Scholarly Open Access*, Accessed June 20, 2016, https://scholarlyoa.com/2016/02/02/bogus-polish-journal-has-completely-fake-editorial-board/. See also Piotr Sorokowski et al., "Predatory Journals Recruit Fake Editor," *Nature*, March 23, 2017.

21. We therefore disagree both with the use of the term "predatory" and with the logic of predation that is actually mobilized in some of the literature on these practices, like Alexander M. Clark and David R. Thompson, "Five (Bad) Reasons to Publish Your Research in Predatory Journals," *JAN* 73(2017):2499–2501.

22. The complexity of the forensic analysis and the proprietary nature of the sources is exemplified by an early case of fake peer reviews, which "sparked a 14-month investigation that came to involve about 20 people from SAGE's editorial, legal and production departments. It showed that the Gmail addresses were each linked to accounts with Thomson Reuters' ScholarOne, a publication-management system used by [various] publishers.… Editors were able to track every paper that the person or people behind these accounts had allegedly written or reviewed.… As they worked through the list, SAGE investigators realized that authors were both reviewing and citing each other at an anomalous rate." Cat Ferguson, Adam Marcus, and Ivan Oransky, "Publishing: The Peer-Review Scam," *Nature*, 515:480–482, doi:10.1038/515480a.

23. Adam Marcus, "Journal Editors Still Don't Like Talking about Misconduct—And That's a Problem," *Retraction Watch*, August 2, 2018, Accessed August 14, 2018, https://retractionwatch.com/2018/08/02/journal-editors-still-dont-like-talking-about-misconduct-and-thats-a-problem/.

References

Aschwandan, Christie. 2015. "Science Isn't Broken." *FiveThirtyEight*. https://fivethirtyeight.com/features/science-isnt-broken/#part4.

Beall, Jeffrey. 2016. "Bogus Polish Journal Has Completely Fake Editorial Board." *Scholarly Open Access*. https://scholarlyoa.com/2016/02/02/bogus-polish-journal-has-completely-fake-editorial-board/.

Beall, Jeffrey. 2016. "Look Out for Bogus Impact Factor Companies." *Scholarly Open Access*. https://scholarlyoa.com/2013/08/06/bogus-impact-factor-companies.

Biagioli, Mario. 2016. "Watch Out for Cheats in Citation Game." *Nature* 535(7611):201. doi:10.1038/535201a.

Biagioli, Mario. 2018. "Quality to Impact, Text to Metadata: Evaluation and Publication in the Age of Metrics." *KNOW* 2(2):1–27.

Biagioli, Mario, Martin Kenney, Ben Martin, and John Walsh. 2019. “Academic Misconduct, Misrepresentation, and Gaming.” *Research Policy* 48: 401–413.

Bik, Elisabeth M., Arturo Casadevall, and Ferric C. Fang. 2016. “The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications.” *bioRxiv*; 049452v2. http://biorxiv.org/lookup/doi/10.1101/049452. May 17, 2016.

Blatt, Michael R. 2015. “Vigilante Science.” *Plant Physiology* 169(2):907–909.

Bohannon, John. 2013. “Who’s Afraid of Peer Review?” *Science* 342(6154):60–65.

Bohannon, John. 2015. “Hoax-Detecting Software Spots Fake Papers” *Science* 348(6230):18–19.

Brooks, Margaret. 2009. “Red-Flag Conferences.” *The Chronicle of Higher Education*. March 26, 2009.

Brunton, Finn. 2013. *Spam: A Shadow History of the Internet.* Cambridge, MA: MIT Press.

Burrows, Roger. 2012. “Living with the H-Index? Metric Assemblages in the Contemporary Academy.” *The Sociological Review* 60(2):355–372.

Campbell, Donald T. 1979 (1975). “Assessing the Impact of Planned Social Change.” *Evaluation and Program Planning* 2(1):67–90. doi:10.1016/0149–7189(79)90048-X.

Clark, Alexander M., and David R. Thompson. 2016. “Five (Bad) Reasons to Publish Your Research in Predatory Journals.” *Journal of Advanced Nursing*. August. doi:10.1111/jan.13090.

Csiszar, Alex. 2018. *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century*. Chicago: University of Chicago Press.

Ferguson, Cat, Adam Marcus, and Ivan Oransky. 2014. “Publishing: The Peer-Review Scam.” *Nature* 515(7528):480–82. doi:10.1038/515480a.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Misuses*. Cambridge, MA: MIT Press.

Grant, Bob. 2016. “US Gov’t Takes On Predatory Publishers.” *The Scientist*. August 29.

Grove, Jack. 2016. “Robot-Written Peer Reviews.” *Inside Higher Ed.*. September 22. https://www.insidehighered.com/news/2016/09/22/many-academics-are-fooled-robot-written-peer-reviews#.V-Qd_s9vpF0.gmail.

Hvistendahl, Mara. 2013. “China’s Publication Bazaar.” *Science* 342(6162):1035. doi:10.1126/science.342.6162.1035.

Jacob, Marie-Andrée. 2014. “Misconduct Hunting: Research integrity via Law Science and Technology,” in E. Cloatre and M. Pickersgill, eds, *Knowledge, Technology and Law*, 137–153. London: Routledge.

Jalalian, Mehrdad. 2015. “The Story of Fake Impact Factor Companies and How We Detected Them.” *Electronic Physician* 7(2):1069–1072. http://doi.org/10.14661/2015.1069–1072.

Jie, Xu, and Matthias Wahls. 2012. “The Scholarly Publishing Industry in China: Overview and Opportunities.” *Learned Publishing* 25:63–74.

Kaiser, Jocelyn. 2016. “New Leader of NIH’s Research Watchdog Faces Staff Revolt.” *Science*. August. doi:10.1126/science.aah7226.

Kutner, Max. 2014. “How to Game the College Rankings.” *Boston Magazine*. August 14.

Lippman, Alexandra, and Christopher Kelty. 2019. “In a Race for Mentions, It’s Open Season on Researchers.” *Nature Index*. January 18. https://www.natureindex.com/news-blog/in-a-race-for-mentions-its-open-season-on-researchers.

Lock, Stephen, and Frank Wells. 2001. *Fraud and Misconduct in Biomedical Research*. Hoboken, NJ: Wiley-Blackwell.

Lucas, R. E. 1976. “Econometric Policy Evaluation: A Critique,” in R. E. Lucas, ed, *Studies in Business-Cycle Theory*, 104–130. Cambridge, MA: MIT Press.

Macdonald, Stuart, and Jacqueline Kam. 2007. “Ring a Ring O? Roses: Quality Journals and Gamesmanship in Management Studies.” *Journal of Management Studies* 44(4):640–655. doi:10.1111/j.1467–6486.2007.00704.x.

Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: University of Chicago Press.

Pain, Elisabeth. 2014. “Paul Brookes: Surviving as an Outed Whistleblower.” *Science*. March. doi:10.1126/science.caredit.a1400061.

Power, Michael. 1997. *Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

PLoS Medicine Editors. 2006. “The Impact Factor Game. It Is Time to Find a Better Way to Assess the Scientific Literature.” *PLoS Med* 3(6):e291.

Raymond, Eric R. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary.* Sebastopol, CA: O’Reilly.

Rossner, Mike, and Kenneth M. Yamada. 2004. “What’s in a Picture? The Temptation of Image Manipulation.” *The Journal of Cell Biology* 166(1):11–15. doi:10.1083/jcb.200406019.

Shao, Jufang, and Huiyun Shen. 2011. “The Outflow of Academic Papers from China: Why Is It happening and Can It Be Stemmed?” *Learned Publishing* 24:95–97.

Shuhua, Wang, and Paul Weldon. 2006. “Chinese Academic Journals: Quality, Issues, and Solutions.” *Learned Publishing* 19:97–105.

Sorokowski, Piotr et al. 2017. “Predatory Journals Recruit Fake Editor.” *Nature*. March 23.

Steneck, Nicholas. 2004. *ORI Introduction to the Responsible Conduct of Research*. Washington, DC: Office for Research Integrity.

Strathern, Marilyn. 1997. “‘Improving Ratings’: Audit in the British University System.” *European Review* 5:305–321.

Strathern, Marilyn, Editor. 2000. *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. London: Routledge.

Straumsheim, Carl. 2016. “Feds Target ‘Predatory’ Publishers.” *Inside Higher Ed.*. August 29. https://www.insidehighered.com/news/2016/08/29/federal-trade-commission-begins-crack-down-predatory-publishers.

Van Noorden, Richard. 2012. "Record Number of Journals Banned for Boosting Impact Factor with Self-Citations." *Nature*. June 29.

Van Noorden, Richard. 2013. "Brazilian Citation Scheme Outed." *Nature*. August 27.

Van Noorden, Richard. 2014. "Publishers Withdraw More Than 120 Gibberish Papers." *Nature*. February 24. http://www.nature.com/news/publishers-withdraw-more-than-120-gibberish-papers-1.14763.

West, Jevin D., Carl T. Bergstrom, and Ted C. Bergstrom. 2010. "The Eigenfactor Metrics: A Network Approach to Assessing Scholarly Journals." *College & Research Libraries* 71(3):236–244.

Wilhite, Allen, and Eric A. Fong. 2012. "Coercive Citation in Academic Publishing." *Science* 335(6068):542–543. doi:10.1126/science.1212540.

I

Beyond and Before Metrics

Before and Beyond Metrics begins with the origins of publication metrics, tracing the contingencies of their genealogy to both question the present and to envisage possible future postmetrics scenarios. Rather than analyzing the manipulation of a specific metric, this section considers the gaming that is necessarily involved in the introduction of any metrics—not the gaming of an established game, but the gaming that goes into defining the game itself. Metrics are not set once and for all, but are rather introduced and modified through a never-ending process propelled by gaming itself. Any metric will create the possibility of its gaming (and gaming-related misconduct), which will eventually crowd that market, thus creating an incentive to modify the metrics, which in turn will usher in the next generation of innovative gaming and manipulations. Metrics appear to set specific targets, but those targets are inexorably bound to be moving ones. Several contributions to this section show how Goodhart's law, despite its obvious value, does not capture the fact that—from faculty performance to university rankings—academic metrics is not one thing, but many different factors, rankings, and indicators jockeying for attention in an increasingly tight market, marginalizing some competitors while forcing others to focus on different indicators and niches. (The remarkable variety—if not outright incommensurability—of international university rankings exemplifies this trend).

Alex Csiszar shows that the problem of gaming metrics, far from being an exogenous pathology, is part and parcel of the history of scientometrics. Well before scientific indicators came into common use, Robert Merton wrote a letter to Eugene Garfield, the owner of the Institute for Scientific Information and the inventor of the Science Citation Index, predicting that a "goal displacement" would inevitably emerge if scientometrics were to be used not to map the dynamics of the scientific community (as Garfield had initially proposed), but to evaluate and reward the publication performance of specific scientists. In a striking foretelling of Goodhart's law, Merton stated the following: "Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did." The future of scientometrics unfolded precisely as Merton had predicted, creating a field that became extraordinarily influential precisely because it was "hacked" and turned from a descriptive into an evaluative discipline, thus spawning a potentially endless range of indicators, and their gaming.

Yves Gingras puts the local history of scientometrics into the context of large-scale global economic changes that have affected scholarly publication to argue that "the internet revolution, the economic transformation of journal publishing and the evaluation turn" created the "perfect storm," which transformed scientific publishing and contributed to the rise in academic misconduct. Gingras's conclusion resonates well with Merton's prediction: the scientific paper has been transformed "from a *knowledge unit* to an *accounting unit* used to evaluate researchers."

Michael Power and Paul Wouters move from a focus on scientific papers to one on institutions. Power argues that the university's demand that its faculty demonstrate the impact of their research "has become a *game*, understood as an infrastructure for the production of a certain kind of truth (Foucault)." In this meta-game, the order of scholarly production has been flipped upside down: researchers must make sure that they will have impact "*before* they do the related research. Rather than impact being an outcome of research, it is research, or a certain style of research, which is becoming the product of the impact apparatus." The need to be found to have had impact determines the kind of research one will do and the questions one will ask.

Paul Wouters expands on the strange alchemy of "impact" by reflecting on how academic research has become "a strategic business in which it is increasingly vital for researchers to be visible at both the national and international level." This need for high visibility has curtailed academic autonomy, subjecting researchers to continuous evaluations and assessments of the scientific excellence and societal impact of their work. Given how research has become a strategic business, Wouters concludes that it has become difficult to distinguish "gaming the system" from "properly functioning in the system."

Sharing many of the concerns of the previous contributors, James Griesemer offers a radical proposal to move us toward a postmetrics future. Griesemer moves beyond the mantric reiteration of Goodhart's law and the a priori critique of metrics it entails, proposing we study metrics as just any other dimension of scientific practice (including instruments, funding, and training) and to do so "experimentally." That is, we should not just look at the effects that metrics has on those who respectfully follow it, but also at the unexpected consequences that happen when one actively "hacks" metrics. Such experimental hackings should be practiced and understood as a form of productive misbehaving (to be clearly distinguished from "behaving badly"): "just as video game hackers improve game play by intentionally violating the designs of the game designers to make the game play differently, science studies research might involve experimental manipulation of metrics as a means of understanding contemporary science in the age of metric tides." And because metrics is about publications and their reception, a journal or series of journals explicitly dedicated to the experimental study of metrics would be the "laboratory" for this kind of research that hacks metrics for knowledge, not for profit.

1

Gaming Metrics Before the Game: Citation and the Bureaucratic Virtuoso

Alex Csiszar

> Henceforth, academic interviewees, it won't be: How many papers have you published? but: Where? … Publish at the *top* or be damned.
> —Correspondence in *New Scientist*, November 23, 1972.

On June 13, 1974, in the foothills of Stanford University, a group of social scientists, information entrepreneurs, and government officials gathered to deliberate what had recently become pressing questions: Can science be measured? How? And what might be the consequences of doing so? The conference on "science indicators" at the Center for Advanced Studies in the Behavioral Sciences (CASBS) was among the first ever held on the subject. But the invitations sent to participants had already raised the question how measuring scientific productivity might alter the very object under study: "the creation and official use of indicators may lead to shifts in the range and kind of scientific activity itself."[1]

Proposals to establish objective measures of individual and national scientific activity have been around for nearly two centuries.[2] But during the late 1960s in the United States, renewed pressure on scientific funding bodies, sociological interest in studying what had become known as "the scientific community," and the increasingly wide availability of citation data put a spotlight on these questions.

It is commonly supposed that rising incidence of gaming is an unintended consequence of these developments, one that an era more innocent than our own simply did not have much reason—or enough foresight—to confront. We have even adapted a term from economics—Goodhart's law—to capture the idea that when measures become standards, they cease to be good measures. By unearthing a set of episodes from the early history of citation analysis, I want to argue two things. First, this supposition is historically mistaken: the idea that attempts to measure scientific output might lead to changes in researcher behavior has been a part of the conversation about citation metrics from the beginning. Second, the very framing of the problem in terms of unintended consequences not only has a long history, but also has caused observers to ask a limited set of questions about gaming and misconduct and thus to produce an equally limited set of responses to it.[3]

Many circumstances had come together to bring about the conference on science indicators. The catalyst was the publication by the National Science Foundation (NSF) of the first edition of *Science Indicators* (hereafter SI-72), a collection of statistics on the relative performance of science in the United States during 1972. In the late 1960s, at the instigation of Congress, the NSF had undergone its first major restructuring since it was established in 1950. The resulting 1968 amendment to the NSF act brought the foundation more closely under the control of Congress and the executive branch (Committee on Labor and Welfare, 1968). In particular, there would now be annual hearings to approve expenditures, and approval would depend on an annual report from the foundation on the "status and health of science and its various disciplines." The legislation did not specify the form of these reports but there were strong hints that numbers were what was wanted. Henceforth, the NSF owed the people hard data about the outputs of science.

The new mandate led NSF officers to the notion that they might construct a comprehensive and standardized set of indicators of the health of science. Congress received the first such report in 1973. The NSF then reached out to the Social Science Research Council—already studying the construction of *social* indicators—about how they had done in this first effort. The latter thought this an ideal job for sociologists of science. Robert K. Merton, the towering figure in the field, got the call.

The timing could not have been better. That academic year, Merton and his collaborator and partner Harriet Zuckerman were on leave at CASBS in Palo Alto, where they had assembled an interdisciplinary working group to pursue projects in the historical sociology of science. The core group included Nobel Prize–winning biologist Joshua Lederberg and the historians Arnold Thackray and Yehuda Elkana. The project on indicators could prove the relevance of the new program in sociology of science, Merton thought. It would be "a test for the guild."[4]

"Fun and Games with Citations"

In a small room uncomfortably packed with scientists and sociologists (Merton explained that "enforced intimacy is more conducive to straight-out serious talk"), one of the most consequential participants had no academic affiliation at all. Eugene Garfield was the proprietor of the Institute for Scientific Information (ISI) and the inventor of the Science Citation Index (SCI). Much of the debate about measuring science that had arisen in the past decade was focused on analyzing citation data obtained (at a price) from his company. Garfield was asked to precirculate a paper on the ways in which citation data could be used to track the health of science, its cognitive development, and its goals. Nearly all the other papers were dedicated to more or less savage critiques of SI-72 from a variety of perspectives, including statistics (William Kruskal), economics (Zvi Griliches), history and philosophy of science (Gerald Holton), and political philosophy (Yaron Ezrahi). The only other constructive paper—by the brothers Jonathan and Stephen Cole, students of Merton—also relied on Garfield's work; it explored how citation analysis could be used to identify scientific consensus.[5] Garfield's invention was under the microscope.

In his role as convener, Merton often steered the conversation. He was especially keen to direct attention to one key problem: "I'd like to get on the record the problem of goal displacement." He explained that "when certain indicators get officially, or quasi-officially, established as measures of this, that, or the other, there will be, one should look for, efforts to manipulate the numbers, by one's behavior":

> In the case of citations, I make a small self-fulfilling prophecy, that has already been fulfilled in one area I know, namely, as more and more citations are used both officially and unofficially as measures of contribution of work of relative standing and the like, there will be diverse patterns of adaptations to those uses, in a feedback way, and citation behavior will change in part in certain sub-sets of the community.

Merton noted, cryptically, that he knew already of a "highly organized effort…to operate on citations for political purposes" and he worried that the "purity of the community of science is itself, maybe, put in jeopardy" by this new behavior.[6]

So important did Merton believe this problem to be that he continued to impress it on Garfield in the months following the conference. "I enjoyed your fun and games with citations," he wrote, perhaps with a hint of mockery. And he warned Garfield to exercise care:

> Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did. Now that more and more scientists and scholars are becoming *acutely* conscious of citations as a form of behavior, some will begin, in semi-organized fashion, to work out citation-arrangements in their own little specialties.

This was as true for science, Merton said, as it was for any bureaucratic system involving counting and rewards: censuses, employee performance measures, or government agencies.[7]

Merton knew all this well, for he had introduced the concept of goal displacement to sociology himself several decades earlier. In 1940, he had written an essay, "Bureaucratic Structure and Personality," which focused on the dysfunctional consequences of bureaucratic organizations. "Goal displacement" was the process by which "an instrumental value becomes a terminal value." This was an important case of the unintended consequences of purposive action; it occurred whenever rules or performance measures were set up, which gradually came to be adhered to by actors as ends in themselves. Over the next decades, many of Merton's students helped found the US field of organizational sociology, studying government agencies, industrial firms, and other organizations to determine how rules and performance metrics can lead to deviant behavior and unintended outcomes—goal displacement, in short.

At the end of the 1950s, after major contributions to several areas of sociology, including a major intervention in structural-functionalism sociological theory, Merton shifted his research focus back to the sociology of science (the topic that he had written for his PhD). In this field, two of his papers had already become classics (Merton, 1938, 1942). These papers outlined a set of behavioral norms that governed the social structure of science. Insofar as these norms—universalism, communism, organized skepticism, and disinterestedness—were adhered to by practitioners, he had argued, science progressed toward greater and more certain knowledge. But this had always been a rather theoretical idea. As he returned to the issue (Merton, 1957), he realized that to observe these norms in action, he needed to get more concrete: just how were they instantiated in actual practice? Publishing practices seemed to be the nerve center where issues of reward, responsibility, and status merged on an everyday basis.

But the value placed on publishing papers, which made "open disclosure" the rule in science, also contained within it the possibility of pathology. As this value became institutionalized, Merton (1969) explained, it increasingly encouraged "through the displacement of goals, a spurious emphasis on publication for its own sake, almost irrespective of the merit of what is published."[8] Scientists published more and more papers, but each of them said less and less.

The same went for other measures of achievement. Merton began to enroll graduate students and to run seminars on the sociology of science in the 1960s, and he was on the lookout for sources of data for research projects. The group explored bibliometric data, biographical data, and survey methods. Through Garfield's efforts to market his products to sociologists and historians, Merton learned of the SCI in 1962. By 1965, Merton's students were mining the citation index for insights on the development of schools and fields, the growth of scientific consensus, and reward systems. By 1970, he was tentatively employing citation analysis in his own work, but already warning of its dangers. In a paper on age stratification in science (Zuckerman and Merton, 1972), a long footnote warned that the growing use of citation analyses "as aids in deciding upon the appointment and promotion of scientists" could lead to changes in citation practices that would invalidate them as measures altogether.

By the time of the CASBS conference in 1974, Merton was far from alone in expressing concern over the role that paper counting and citation counting might do to science. The conference had come at a critical moment for Garfield's enterprise. During the 1960s, he had worked diligently to market the SCI to scientists as an aid to information retrieval (Wouters, 1999). Looking to explore other markets for his expensive tools, he had also reached out tentatively to sociologists and others concerned with studying science, asking politely whether they might find some use in it. But he tended, in public, to play down suggestions that the citation index might be used as a means of evaluating scientists or research programs.

A decade later, however, Garfield had become bolder, and he mounted a public campaign to argue for new uses of citation analysis. In 1970, he wrote in *Nature* that "increasingly scarce intellectual and financial resources for supporting research could be managed more efficiently" using citation counting to identify the most promising and creative researchers. In 1972, he published in *Science* his first public ranking of journals according to the frequency with which they came up in citations—both absolutely and relative to the number of articles published. That year he also began marketing *Journal Citation Reports*, which became the permanent home of the "Impact Factor."

Garfield's early steps toward legitimating the evaluative use of citation analysis—along with other investigations by Derek J. de Solla Price (1969), Joel Margolis (1967), and the Cole brothers (1972)—made it a topic of controversy in scientific circles. Many observers were appalled, arguing that such surveillance and the encouragement of hierarchy were inimical to the spirit of creativity and equality central to the scientific community. Garfield's ranked list of journals was "a pedantic dunghill," suggested one letter in *New Scientist*, a "highly invidious pecking order" calculated to appeal to "connoisseurs of hierarchies."[9] Others, such as Samuel Goudsmit (1974), legendary editor of the *Physical Review*, argued that the implicit theory of citation of these studies was mistaken, for acknowledgement was only a minor use of footnotes. He dismissed citation analysis and quoted approvingly a warning that equating "high frequency of citation with worth or excellence will end in disaster."

Many critiques were explicit about the potential for gaming. Journal rankings were the equivalent of "pop charts for science," suggested Alex Comfort (1970) in *Nature*. The corrupt music industry gave scientists a taste of "the abuses we might expect (and even … indulge in): Commercial concerns dedicated to lobbying, so that our rating might be raised? Promotional quotation data on the relative one-upmanship of particular journals? In California nothing is impossible." And the mathematician and historian Kenneth O. May (1967) noted that these new uses of citations would lead authors to cite "their friends' papers more (a friend is someone who cites in return)" and also to increase their citations to marginally relevant papers "so as to attract people who might (and perhaps should) miss the paper."

Thus, while Garfield went into the conference with optimism, he also knew that these new uses for citation analysis were coming under attack. But that Merton himself should challenge their legitimacy must have been jarring. Merton had become a great supporter of Garfield and the ISI; he sat on its board, proselytized the uses of the citation index to others, and gave Garfield strategic advice when problems arose. Garfield nevertheless took Merton's warning in stride. "If it should turn out that over the next 10 to 20 years goal displacement does occur with the Science Citation Index," Garfield admitted, "then we will simply have to abandon its uses as an indicator of quality, etc." But he did not really think that this was likely. Maybe editors could be taught "to look out for self-serving attempts to distort citation data?"[10] Merton felt able to reassure Garfield that while localized citation misconduct was sure to happen, it was doubtful that things would get so bad that science itself would really be transformed "for a long, long time, if at all. It would follow upon a massive use of citation counts as a part of the reward system of science: appointments, promotions, research grants, elections to office in professional societies, awards, etc. And that is not in the cards."[11]

Merton was wrong. In the next decades, citation analysis and the Journal Impact Factor (JIF), which became enshrined in *Journal Citation Reports*, spread far and wide. The field of scientometrics blossomed. By 1980, Garfield was considering ways to capitalize on the growing use of citation data in evaluations of individuals and institutions. After several years of publishing Impact Factor data, in 1981, the ISI formulated a plan for a new suite of products marketed for the "deliberate use of citation counts for evaluative purposes." The proposed Science Citation Analysis System (SCAS) would be an online tool for "administrators of research faculty/staff" to measure the productivity of individuals and institutions without relying on such crude markers as the JIF.[12] Again, Garfield and Merton engaged in intense conversation about the potential consequences of such a move for science. Were Garfield's tools simply making more efficient a set of evaluative practices that were already flourishing? Or would the availability of such tools encourage these practices and thus lead to their misuse by administrators and deviant behaviour by those subject to the measures? Merton assured Garfield that the latter would be the case, and that there were no moral grounds on which he could disavow his role in any such an outcome: "You're helping to build the practice which will include a lot of pollution, this misuse which you can't control, but you've been instrumental in making possible."[13]

Even as Merton continued to warn Garfield about goal displacement behind the scenes, he remained a strong supporter of the ISI in public. The term that has come most often to be used to describe the susceptibility to manipulation of metrics of science was not "goal displacement" but rather Goodhart's law.[14] Although the latter originated in monetary economics, the common formulation in this context comes from Marilyn Strathern (1996): "When a measure becomes a target, it ceases to be a good measure." While "goal displacement" describes a similar phenomenon, the emphasis is inverted: where Goodhart's law highlights the consequences for the validity of the metric, Merton's concept highlighted the consequences for the object being measured. This emphasis was especially clear in the studies of bureaucratic organizations pursued by Merton's students. In many cases, measures and rewards that were instituted to encourage behavior that would further the goals of an organization led not simply to dysfunctional behavior, but also to reshaping the perceived goals of the organizations themselves. In the big picture, the phenomenon of gaming matters not simply—or even principally—because it invalidates any given metric, but because it can profoundly shape the everyday conditions under which research is carried out, not simply in terms of how and when people publish their findings, but of what kinds of projects they choose to pursue and how they go about them. Ironically, Strathern's now-famous formulation of Goodhart's law was derived from a paper by Keith Hoskin (1996), which was itself a critique of this formulation along these lines. Hoskin pointed out that the law seems to imply that the target is the problem, and that measures are not in themselves potentially problematic, thus leading to a search for "either better targets, or better ways of anticipating and handling a presumed natural-born predilection for beating the target system." (This is a reasonable characterization, incidentally, of a key argument made by advocates of altmetrics.[15])

More generally, however, the ideas behind goal displacement and Goodhart's law have both constrained the kinds of questions we ask about metrics and their problems. By adapting the theory of bureaucracy to the study of science, Merton encouraged the idea that the social study of science could be pursued along broadly similar lines to the social study of corporations and firms, and in relative isolation from the rest of society. But this obscured a fundamental disanalogy. In a democratic state, there is little chance of there being widespread consensus about what the goals of science ought to be in the first place. Broad public views about the legitimate ends of scientific activity cannot in general be separated from public representations of scientific activity and its productivity. For that reason, public indicators of science affect not only the practice of research, but also the public legitimacy of science.[16] To put it another way, understanding metrics of science is not just about the internal sociology of research, but also about the politics of scientific knowledge (Ezrahi, 1978; 1990).

My sense is that many of the most common critiques of the implementation and use of scientometrics essentially ignore this consideration. They tend to suppose that there is some pure and efficient form of scientific practice—and of scientific publishing—that was once directed wholly to the production of knowledge, but which has recently been turned upside down (for good or ill) by the implementation of accounting practices.[17] The late-Mertonian mode in the sociology of science was at its peak in the 1960s and 1970s, but it remained alive among those experts producing increasingly sophisticated tools with which to measure scientific productivity, even as they grew increasingly frustrated that these tools were subject to misuse and abuse.

When we debate the efficacy of a given measure of scientific achievement, we are also debating (perhaps implicitly) the grounds of the public legitimacy of science itself.[18] This perspective implies that the analysis of metrics in use ought to attend to the political values and modalities of power within which they are constructed, deployed, and spread (Merry, 2016). Metrics have become mediating devices in the interaction between the groups and institutions involved in science policy, its administration, and research. Scientific publishing—whose ascendancy has often been linked to public representations of the legitimacy of modern science—is now front and center in public debate about inefficiency and corruption (Sarewitz, 2016; Oransky, this volume, chapter 10).

Previous histories of measuring science (Godin, 2005; Bellis, 2009; Gingras, 2016) have not generally paid much attention to the question of misconduct and gaming, but we need a new history that puts these matters at the heart of the story. This is not simply because I think we need to get the history right, but because producing a better historical narrative will help to break the spell of Goodhart's law and its several variations. It is not yet clear what the history of scientific metrics *in use* might look like. But I suspect that we will need to formulate Goodhart's counter-law to describe the more common historical sequence: "*only when a measure becomes a target is it widely taken up as a measure worth using*."

Notes

1. See the rough transcript of R. K. Merton's remarks, and the generic conference invitation template dated March 21, 1974, in Robert K. Merton Papers, Columbia University Rare Book & Manuscript Library [hereafter RKMP], 328.4.

2. For an early example of measuring scientific output, see Granville (1830).

3. I am inspired here in part by James Griesemer's essay in this volume.

4. Quoted in Robert Parke letter to R. K. Merton, December 21, 1973, RKMP 328.7.

5. Conference agenda, RKMP 328.4.

6. Rough transcript of R. K. Merton's remarks, RKMP 328.4.

7. R. K. Merton letter to E. Garfield, August 3, 1974, RKMP 168.5.

8. See also, earlier, Merton (1957) and, later, Zuckerman and Merton (1971), in which the authors also make the fascinating hypothesis that the spread of referee systems was a further unintended consequence of the impulse to publish as an end in itself.

9. "At the Heart of the Paper Blizzard," *New Scientist*, November 23, 1972, p. 464. On resistance to "the moral equivalence of the scientist," see Shapin (2008).

10. E. Garfield letter to R. K. Merton, August 19, 1974, and R. K. Merton's reply, September 2, 1974, RKMP 168.5.

11. R. K. Merton to E. Garfield, September 2, 1974, RKMP 168.5.

12. SCAS is described in L. Simon, "Preliminary Specification for Science Citation Analysis System (SCAS)," January 29, 1981, in RKMP 168.4.

13. Transcript of a recorded conversation between E. Garfield and R. K. Merton on May 23, 1981, Papers of Eugene Garfield, Science History Institute, Philadelphia, 44.5.

14. This is not to say that "goal displacement" has wholly disappeared from the conversation. See, for example, Hicks et al. (2015).

15. See Jennifer Lin in this volume, chapter 16.

16. These points were in fact made at the 1974 conference by the political theorist Yaron Ezrahi (1978). But they seem largely to have been ignored.

17. See, for example, Yves Gingras in this volume, chapter 2. For the historical counterargument, see Csiszar (2018).

18. A similar point is made at the end of Paul Wouters's contribution to this volume, chapter 4.

References

Bellis, Nicola De. 2009. *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics*. Lanham, MD: The Scarecrow Press.

Cole, Jonathan R. 2000. "A Short History of the Use of Citations as a Measure of the Impact of Scientific and Scholarly Work." In: Cronin, B., Atkins, H. B. (Eds.), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Information Today, Medford, NJ, pp. 281–300.

Cole, Jonathan R., and Stephen Cole. 1972. "The Ortega Hypothesis." *Science* 178:368–375.

Comfort, Alex. 1970. "Pop Charts for Science." *Nature* 227:1069.

Committee on Labor and Welfare. 1968. National Science Foundation Act Amendments of 1968: Hearings Before the United States Senate Committee on Labor and Public Welfare, Special Subcommittee on Science, 90th Congress, First Session, on Nov. 15 and 16, 1967. Washington, DC: US GPO.

Csiszar, Alex. 2018. *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century*. Chicago: University of Chicago Press.

Elkana, Yehuda, et al., ed. 1978. *Toward a Metric of Science: The Advent of Science Indicators*. New York: Wiley.

Ezrahi, Yaron. 1978. "Political Contexts of Science Indicators." In: Elkana, Y., Lederberg, J., Merton, R. K., Thackray, A., Zuckerman, H. (Eds.), *Toward a Metric of Science: The Advent of Science Indicators*. Wiley, New York, pp. 284–327.

Ezrahi, Yaron. 1990. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy*. Cambridge, MA: Harvard University Press.

Garfield, Eugene. 1970. “Citation Indexing for Studying Science.” *Nature* 227: 669–671.

Garfield, Eugene. 1972. “Citation Analysis as a Tool in Journal Evaluation.” *Science* 178:471–479.

Garfield, Eugene. 1973. “Citation and Distinction.” *Nature* 242:485.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation*. Cambridge, MA: MIT Press.

Godin, Benoît. 2005. *Measurement and Statistics on Science and Technology: 1920 to the Present*. London: Routledge.

Goudsmit, Samuel A. 1974. “Citation Analysis.” *Science* 183(4120):28–33.

Granville, Augustus Bozzi. 1830. *Science Without a Head*. London: T. Ridgway.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. “The Leiden Manifesto for Research Metrics.” *Nature* 520:429–431.

Hoskin, Keith. 1996. “The ‘Awful Idea of Accountability’: Inscribing People into the Measurement of Objects.” In: Munro, R., Mouritsen, J. (Eds.), *Accountability, Power, Ethos and the Technologies of Managing*. International Thomson Business Press, London, pp. 265–282.

Margolis, Joel. 1967. “Citation Indexing and Evaluation of Scientific Papers.” *Science* 155:1213–1219.

May, Kenneth O. 1967. “Abuses of Citation Indexing.” *Science* 156 (3777): 890–892.

Merry, Sally Engle. 2016. *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: University of Chicago Press.

Merton, Robert K. 1938. “Science and the Social Order.” *Philosophy of Science* 5:321–337.

Merton, Robert K. 1940. “Bureaucratic Structure and Personality.” *Social Forces* 18(4):560–568.

Merton, Robert K. 1942. “Science and Technology in a Democratic Order.” *Journal of Legal and Political Sociology* 1:115–26.

Merton, Robert K. 1957. “Priorities in Scientific Discovery.” *American Sociological Review* 22:635–659.

Merton, Robert K. 1969. “Behavior Patterns of Scientists.” *American Scientist* 57(1):1–23.

National Science Board, National Science Foundation. 1973. *Science Indicators 1972*. Washington, DC: US GPO.

Price, Derek J. de Solla. 1969. “Measuring the Size of Science.” *Proceedings of the Israel Academy of Sciences and Humanities* 4(6):98–111.

Sarewitz, Daniel. 2016. “Saving Science.” *New Atlantis* 49:4–40.

Shapin, Steven. 2008. *The Scientific Life: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press.

Strathern, Marilyn. 1996. “From Improvement to Enhancement: An Anthropological Comment on the Audit Culture.” *Cambridge Anthropology* 19(3):1–21.

Wade, Nicholas. 1975. “Citation Analysis: A New Tool for Science Administrators.” *Science* 188:429–432.

Wouters, Paul. 1999. “The Creation of the Science Citation Index.” In: Bowden, M. E., Hahn, T. B. (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems*. Information Today, Medford, NJ, pp. 127–136.

Zuckerman, Harriet, and Robert. K. Merton. 1971. “Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System.” *Minerva* 9(1):66–100.

Zuckerman, Harriet, and Robert K. Merton. 1972. “Age, Aging, and Age Structure in Science.” In: *Aging and Society*. Russell Sage Foundation, New York, pp. 292–356.

2

The Transformation of the Scientific Paper: From Knowledge to Accounting Unit

Yves Gingras

Since the mid-1990s, observers and actors in the scientific field—as defined in Bourdieu (1986, 2004) as a structured space of agents and institutions in competition for the accumulation of credit or "symbolic capital"—have commented on the many facets of an ongoing major transformation in the structural conditions of scientific practice: massification of research, mounting pressure to publish, relative decline of government investments, and the arrival into the research system of the ideology of "knowledge management" with its insistence on quantitative evaluation measures of productivity and "impact" of academic research (Bruneau and Savage, 2002). As discussed by many contributions to this book, these transformations of the research system led actors to respond with various strategies of "gaming" the system through manipulating the metrics to attain the required results or, as suggested by Griesemer in his contribution, to try to derail the whole enterprise of metrics, though it is not clear how one could do that. By the end of the twentieth century, the technical infrastructure of journal publishing had also started to be radically transformed through the use of internet and electronic publishing (Thompson, 2005). These new technologies of communications made it possible to skip the materiality of the scientific paper and directly produce digital papers and journals that could then circulate much faster and globally on the internet, in turn contributing to transforming the dynamic of scientific practice. Finally, the mounting concentration of scientific journals in the hands of a limited number of giant publishing firms submitted to an increasing demand of profitability on the stock exchange, engendered in the mid-1990s what has been called a "crisis in scholarly publishing" due to the mounting price of journal subscription for academic libraries (Thatcher, 1995; Tenopir and King, 1997; McGuigan and Russell, 2008; Larivière, Haustein, and Mongeon, 2015). An answer to that crisis has been the emergence of the "open-access" movement (Laakso et al., 2011).

The combination of internet technology and the profitability of academic journals, which could now be born digitally, gave rise to the multiplication of specialist journals, many considered of dubious quality, trying to capture a part of the value generated in answering the "offer" of more papers being produced by researchers answering the "publish or perish" injunction from their institutions. Finally, internet diffusion of papers in open access made results more easily accessible to new nonspecialist audiences who could in return influence choices and priorities of research.

Each of these three major processes (the internet revolution, the economic transformation of journal publishing, and the evaluation turn) have generated specific effects. Many of them are analyzed in this book, but my thesis is that it is *their mutual interactions* in a kind of "perfect storm" that made possible the radical changes observed in the scientific field, which saw the rise of deviant behavior on the part of scientists and journal editors as well as managers of academic institutions. One may think here of the multiplication of "predatory journals," faked peer review, the manipulation of citations to raise impact factors of journals and positions in academic rankings, and the rise in the number of "corrections" to and "retraction" of scientific papers (Corbyn, 2009; Van Noorden, 2013; Haug, 2015; Beal, 2016), topics discussed below in the contributions collected in sections 2 and 3 of this book.

In this chapter, I propose a global macro-structural analysis that aims at showing how these different transformations have become connected as a consequence of the transformation of the paper from a *knowledge unit* to an *accounting unit* used to evaluate researchers and research organizations (departments, laboratories, and universities). The present structure of the scientific field is thus the result of the complex interrelations of different causal series that converged to change the social function of the scientific paper, making it fit with the twin ideologies of the New Knowledge Management and the Audit Society (Power, 2000; Wilson, 2002). In order to better understand the nature of these recent transformations, I must first present the basic social dynamic that propelled scientific research for more than three centuries. I limit myself to academic research and publications in academic journals and exclude research done for profit or to obtain patents, as this dynamic—though obviously more and more connected to that of the scientific field in response to declining government investments in "basic" research—obeys a different logic, namely that of profit. As Sismondo's contribution to this book shows, the production of papers by pharmaceutical companies is not an end in itself but only a means to market their products (pills and syrups…). By contrast, scientists' final product in the scientific field is the paper, supposed to contain new knowledge, hence my term "unit of knowledge."

The Classic Cycle of Accumulation of Symbolic Capital

The classical sociology of science developed by Robert K. Merton and his school has well established that symbolic recognition (or credit) lies at the basis of the dynamic of scientific research (Hagstrom, 1965; Cole and Cole, 1973; Merton, 1973). In exchange for making known to the scientific community, through publication, a "discovery" (new fact, instrument, interpretation or theory), the scientist obtains the symbolic recognition of his original work. Such a social recognition then gives better chances to access material and economic resources and institutional positions to make further research. This simple but powerful model has been extended by Pierre Bourdieu to take more explicitly into account the links and mutual transformation between recognition, now seen as a form of *symbolic capital,* and the other kinds of capital (economic, cultural, social) that are also at play in the scientific field (Bourdieu, 1986, 2004).

Figure 2.1 shows how that cycle of credibility, to use Latour and Woolgar's reformulation of Bourdieu's ideas, works (Latour and Woolgar, 1982). To simplify the figure, I exclude the cases when papers are rejected. In fact, the model predicts that too many rejections, which means lack of recognition by the community, will lead the affected scientist to abandon research and search for a *different kind* of recognition, often in a different social field. This cycle is premised on the fundamental but implicit idea that the scientific paper is the embodiment of what we can call a "unit of knowledge," that is any original contribution that is perceived as legitimate by the members of a particular scientific field. Generally, the "unit of knowledge" comes only after being peer reviewed and published in a journal, as its value and legitimacy comes from the social recognition and is not intrinsic to the submitted article. Depending on the value (large or small) of that unit of knowledge, which in some fields could even be lines of computer codes, the author then accrues more or less symbolic capital, in particular through being cited by other researchers. The citations accrued in the scientific field thus work as a *symbolic credit allocation mechanism*. The accumulation of symbolic capital facilitates, in turn, access to new resources (including grants, research assistants, and postdocs) for research and favorable institutional positions. With better resources, one can then have better instruments and thus raise the probability of producing better papers and larger "units of knowledge."

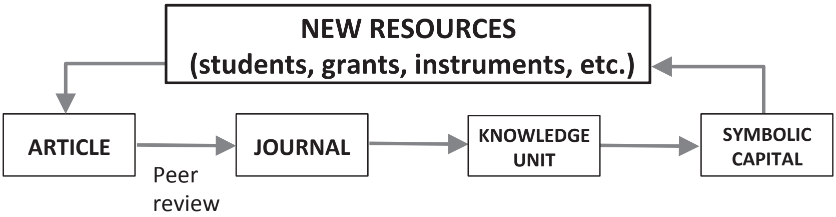


Figure 2.1
The classic cycle of knowledge production, evaluation, and accumulation of credit.

In this system, the obvious crucial step is "peer review," for it is at this very point in the process that the paper will exist or not as a socially recognized unit of knowledge circulating in the scientific field. Since its institutionalization in the middle of the seventeenth century, it has of course evolved in its detail (blind review or not, external or done by the editor), but retains its basic function of gate-keeping the entry into the scientific field (Zuckerman and Merton, 1971; Chubin and Hackett, 1990; Biagioli, 2002; Kronick, 2004; Pyenson, 2008). It could thus be expected that the mounting pressure on the research system since the mid-1970s would generate a mounting critique of peer review (Gustafson, 1975). By the end of the 1980s, the peer-review mechanism even became the object of systematic and continued analysis, particularly in the biomedical sciences, through an annual congress of peer review (Rennie, 2016). But despite all the critiques addressed to peer review, it remains central to the basic cycle of accumulation of scientific credit. When the scientific paper will have been transformed into a simple accounting unit used for the evaluation of researchers and their institutions, we will observe the emergence and growth of various more or less deviant behaviors trying to cope with the multiplications of metrics through manipulating them or even the peer-review process itself.

The Multifarious Effects of Research Evaluation

Though bibliometrics developed as a research specialty in the 1960s and 1970s studying the aggregate properties of publications and citations for the use of librarians and historians and sociologists of science, its methods became increasingly used, especially since the 1990s, to facilitate the evaluation of researchers (Gingras, 2016; Wouters's contribution to this book, chapter 4). This development was in line with the ideology of benchmarking and "knowledge management," which invaded the public sector in the 1980s and the universities in the 1990s (Bruno and Didier, 2013). As the tradition of qualitative peer review for promotion and tenure became more and more criticized in the mid-1970s for being “subjective,” quantitative indicators based on counting papers started to be used as a measure of the “productivity” of researchers (Wade, 1975). Bibliometric methods thus provided the basic measures of “impact” used by promoters of “benchmarking” and “league tables” and other “rankings” supposed to evaluate researchers and journals as well as universities (Espeland and Sauder, 2016; Kehm’s contribution to this book, chapter 6).

The move to digital versions of journals and papers would not in itself have transformed the traditional cycle of credit, for it just made papers accessible through the internet for those who were affiliated to institutions that subscribed to the journals. However, the pressure to publish “accounting units” for evaluation purposes triggered a series of important transformations, synthetized in figure 2.2, that gave rise to new actors and new technologies of measurement. An initial important change, though largely invisible to the scientists, is the very content of bibliometric databases. While the oldest of them, the Science Citation Index (SCI; now the Web of Science), at first coded only information on the first author of a paper (since that was sufficient to retrieve it in an information system), bibliometric databases now include the complete list of all authors of papers, because these data are now used for evaluation purposes and not only for making bibliographic searches. From the 1980s to the 2000s, the only direct measure of the “impact” of scientific papers was the number of citations by other papers as registered in the SCI database, which had a monopoly until the creation of the competing Scopus database by Elsevier in 2004.

A second effect of the new evaluation context, one visible to any active scientist, is that it offered an opportunity for companies to create new journals that could accept those papers. The multiplication of papers, related to the growth of the research system and the added competition for scarce resources, called in the creation of new journals. The creation of new journals has of course always been a part of the growth and diversification of research specialties since the emergence of the first journals in the 1660s (Peiffer and Vittu, 2008; Csiszar, 2018). But what is new is that the combined effect of internet publication and the emergence of “open-access journals,” which transfer the cost from subscribing libraries to authors, has opened the door to what many scientists called “predatory journals,” which, without the need of a large investment in infrastructure (as it was necessary with printed journals), could offer rapid publication to scientists struggling to find outlets for their research results (Grant, 2009a, 2009b; Beal, 2016). As for already well-established journals having a strong brand, like *Nature*, they could maximize their monetary value by creating sister journals, using their main brand as attractor. Thus, after having capitalized for more than a century off the value of a single multidisciplinary journal called *Nature*, the company understood that it could extract even more value from this brand by multiplying the titles associated with it. Since the 2000s, it has thus created more than thirty titles associated with the name *Nature*: *Nature Physics*, *Nature Chemistry*, and so on, betting on the aura effect of the original journal to attract the "best papers" from authors in search of maximum visibility for their research results. Property of the American Association for the Advancement of Science (AAAS), the journal *Science* has also adopted that strategy by creating a series of sister journals like *Science Robotics* and *Science Immunology*, for example.

Now, much as the number of "citations" had become the "objective" measure of the value of papers too numerous to be read carefully by evaluation committees, so the Journal Impact Factor has become the measure of the value of journals too numerous to be known to researchers and their evaluators. First published in 1975 as a measure of the visibility of a journal in the scientific field, it was aimed at librarians as a tool of collection management. Significantly, it only became an object of contention in the scientific field in the mid-1990s, when it had been transformed into a mechanical tool for evaluating *researchers* instead of *journals*. The main drawback of using "citations" to papers as a quantitative indicator of their "value" is that they take time to accrue, whereas "evaluation" thrives on "timeliness." Before citing a paper, one must first read it (or at least see a reference to it), then write a new paper citing it, make it pass through peer review, and then see it published. That means that, generally, it takes at least a few years to begin to be cited and about five to seven years (depending on the field) to reach a peak before observing declining citations as the knowledge becomes standard, what Merton called "obliteration by incorporation" (Merton, 1988; McCain, 2014). Faced with this problematic time delay, evaluators thus tended to replace citations by the more readily accessible Journal Impact Factor. Though it characterizes the *journal* and not the *paper*, it provides an immediate "proxy" measure of "quality" and "impact" that obviates the problem of having to wait a couple of years to know whether or not a given paper has had a scientific impact on the field.

Having become a measure of "quality," the impact factor was then used by companies to promote their journals and attract new authors. This apparently inconsequential change gave journals an added centrality in the evaluation system, as it was access to them that was now measured and valued and no more the content of the paper itself and its visibility (measured by its citations). Hence government officials and research institutions in a number of countries (Pakistan, China, South Korea, and Japan) have even established financial incentives based directly on the numerical value of a journal's impact factor, despite the obvious fact that these values cannot be compared across disciplines (Fuyuno and Cyranoski, 2006). This, in turn, put added pressure on scientists, some of whom went so far as to manipulate the peer-review system of journals to make sure their papers get published (see also Brunton's contribution in this book, chapter 18). Once uncovered, the scam led journal editors to retract the papers on the basis that they had been fraudulently peer reviewed by their authors (Barbash, 2015; Haug, 2015; see also Barbour and Stell's contribution in this book, chapter 11). Hence, these new metrics can directly influence access to the resources needed for research (figure 2.2).

The simplistic nature of metrics based on Journal Impact Factors was much denounced by scientists and bibliometricians. After more than a decade of harsh critiques, many journals, including *Nature* and other "elite journals," took some distance from an indicator they had long used to brand themselves, in order not to be associated with simplistic evaluation methods that could tarnish their reputation (Callaway, 2016). As journals supposed to publish units of knowledge, they thus seem to resist being transformed into a mere measurement tool. A similar situation happened with academic book publishers in the 1990s. Pressured by authors to publish the book they needed to justify tenure, publishers asked universities to decouple the tenure system from university presses (Waters, 2004).

The central place now accorded to quantitative indicators of productivity and impact also created a new market for companies offering their service to "evaluate" researchers, laboratories, and universities (American Association of University Professors, 2016). As papers and journals were being transformed into accounting units, journal publishing companies saw a new opportunity to market themselves as providing new evaluation indicators based on their own set of journals, while at the same time enrolling institutions to push their researchers to publish in their self-proclaimed "high-quality" journals. The best example of such a strategy is the creation in 2014 of the "Nature Index," which ranks countries and institutions on the basis of the papers they publish in only sixty-eight journals, defined as "high-quality science journals," and which include seventeen (25 percent) from the *Nature* group itself, owned by Macmillan and now merged with Springer, one of the largest publishers of scientific journals. This new ranking is said to offer "a perspective on high-quality research on the basis of published articles" and to "provide institutions with an easy means to identify and highlight some of their best scientific research."[1]

In addition to radically transforming the mode of circulation of scientific papers, internet access also affected the kind of metrics promoted to measure the "impact" (rarely clearly defined) of research. Through the internet, one can access the article and look at the screen for a minute or download a copy of the entire paper. These actions leave traces and thus create new data on the "uses" of scientific papers. This situation gave rise to "altmetrics," presented as an *alternative* to citation analysis, though they simply measure something different than citations (Priem et al., 2010; see also Lin's contribution to this book, chapter 16). Whereas citations are essentially coming from within the scientific field, measures of internet uses survey a larger spectrum of users of unknown nature in the larger social space, as the "clicks" don't tell us if the author is a simple citizen or a publishing researcher.

By broadening the spectrum of metrics, internet-based measures of the visibility of publications have also contributed to the rush to measure the "impact" of scientific research, even though the exact meaning of those "impacts" remains evanescent. The evangelists of these new web measures have been quick to point out that it takes years to be cited, whereas "views" and "downloads" are accessible in real time, day after day, even hour after hour. By adding blogs and tweets to the array of metrics, one certainly gets tools to follow the diffusion of research results among diverse communities, but that does not say much about the quality or even robustness of the published results. Moreover, these various metrics have radically different temporalities and correspond to different audiences: whereas the scientific field with its citation culture works on the scale of many years, blogs and tweets operate in the larger social space and fluctuate and vanish within days for blogs and within hours for tweets. One can thus expect that scientists will be very critical of institutions trying to use such crude and ill-defined metrics to evaluate the "impact" of their research, be it social, economic, cultural, or whatever else it could be. Pressured to constantly show the usefulness of scientific research, laboratories and universities may nonetheless succumb to the cynical use of these metrics when they suit their purpose, much as they do with the so-called "world university rankings," despite their obvious flaws (Gingras, 2016; see also Kehm's contribution to this book).

Conclusion

After having been in a relatively stable state for more than three centuries, and this despite its continuous exponential growth during that period (Price, 1963), the basic structure of the scientific field has been submitted over the last twenty years to a series of radical transformations. It is not really possible at this point in time to evaluate their long-term effects. We are still in a transition period that sees many kinds of experimentation going on with various ways of making knowledge claims circulate: preprint servers, open reviews, post-publication comments, "fast-track" peer review, and many others. But as the physicists have shown through their use of the preprint server system ArXiv since the beginning of the 1990s, not all disciplines can follow the same model. The potentially dangerous social impact of false medical discoveries announced without peer review would be far more important, for example, than the surprising announcement by physicists of their having observed a neutrino moving faster than the speed of light (Cho, 2012). Though erroneous, this spectacular news item kept the social media and the blogosphere excited for a few days in the fall of 2011, but this major mistake could not have had any serious effect on the daily life of citizens.

One thing seems certain: the pressure to get faster results—and faster evaluation of results—has given rise to a reaction among scientists and the creation of the "slow science" movement (Berg and Seeber, 2016). The manifesto proclaims that "slow science was pretty much the only science conceivable for hundreds of years." The partisans of the movement insist that "society should give scientists the time they need, but more importantly, scientists must take their time […] to think, […] to digest […] to misunderstand each other, especially when fostering lost dialogue between humanities and natural sciences."[2] And only time will tell whether scientists, by becoming more conscious of the mechanisms that triggered the transformations that now affect their research activities, will also take the time needed to make sure the evaluation methods that contributed to transforming the product of their research into mere accounting units get replaced by ones that are more consistent with the nature of research, which, as the manifesto also underlines, "develops unsteadily, with jerky

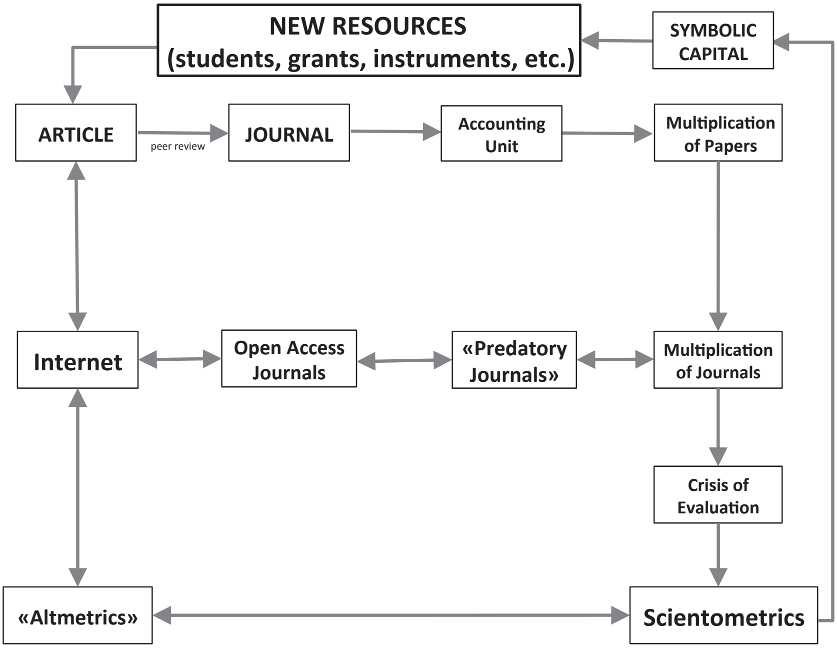

NEW RESOURCES
(students, grants, instruments, etc.)
SYMBOLIC CAPITAL
ARTICLE
peer review
JOURNAL
Accounting Unit
Multiplication of Papers
Internet
Open Access Journals
«Predatory Journals»
Multiplication of Journals
Crisis of Evaluation
«Altmetrics»
Scientometrics

Figure 2.2
How the classic cycle of evaluation (figure 2.1) has been transformed by making the scientific paper a simple accounting unit of research activity, thus generating a series of intermediaries and new evaluation metrics.

moves and unpredictable leaps forward—at the same time, however, it creeps about on a very slow time scale, for which there must be room and to which justice must be done."

Notes

1. http://www.natureindex.com/faq#introduction2. Accessed May 2, 2017.
2. http://slow-science.org/slow-science-manifesto.pdf.

References

American Association of University Professors. 2016. "Statement on 'Academic Analytics' and Research Metrics." March 22: https://www.aaup.org/file/AcademicAnalytics_statement.pdf.

Barbash, F. 2015. "Major Publisher Retracts 43 Scientific Papers Amid Wider Fake Peer-Review Scandal." *Washington Post* (March 27): https://www.washingtonpost.com/news/morning-mix/wp/2015/03/27/fabricated-peer-reviews-prompt

-scientific-journal-to-retract-43-papers-systematic-scheme-may-affect-other-journals/.

Beal, J. 2016. “Predatory Journals: Ban Predators from the Scientific Record.” *Nature* 534(June 16):doi:10.1038/534326a.

Berg, M., and B. Seeber. 2016. *The Slow Professor: Challenging the Culture of Speed in the Academy*. Toronto: University of Toronto Press.

Biagioli, M. 2002. “From Book Censorship to Academic Peer Review.” *Emergences* 12(1):11–45.

Bourdieu, P. 1986. “Forms of Capital,” in *Handbook of Theory of Research for the Sociology of Education*, edited by J. E. Richardson, 241–258. New York: Greenwood Press.

Bourdieu, P. 2004. *Science of Science and Reflexivity*. Chicago: University of Chicago Press.

Bruneau, W., and C. D. Savage. 2002. *Counting Out the Scholars: How Performance Indicators Undermine Universities and Colleges*. Toronto: James Lorimer.

Bruno, I., and E. Didier. 2013. *Benchmarking: L’État Sous Pression Statistique*. Paris: La Découverte.

Callaway, E. 2016. “Publishing Elite Turns Against Impact Factor.” *Nature* 535(July 14):210–211.

Cho, A. 2012. “Once Again, Physicists Debunk Faster-Than-Light Neutrinos.” *Science* (June 8): http://www.sciencemag.org/news/2012/06/once-again-physicists-debunk-faster-light-neutrinos.

Cole, J. R., and S. Cole. 1973. *Social Stratification of Science*. Chicago: University of Chicago Press.

Corbyn, Z. 2009. “Retractions Up Tenfold.” *Times Higher Education* (August 20): https://www.timeshighereducation.com/news/retractions-up-tenfold/407838.article.

Chubin, D., and E. Hackett. 1990. *Peerless Science. Peer Review and US Science Policy*. Albany, NY: SUNY Press.

Espeland, N. W., and M. Sauder. 2016. *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.

Fuyuno, I., and D. Cyranoski. 2006. “Cash for Papers: Putting a Premium on Publication.” *Nature* 441(June 15):792.

Gingras, Y. 2016. *Bibliometrics and Research Evaluation. Uses and Abuses*. Cambridge, MA: MIT Press.

Grant, B. 2009a. “Merck Published Fake Journal.” *TheScientist.com* (April 30): http:// www.the-scientist.com/blog/display/55671/.

Grant, B. 2009b. “Elsevier Published 6 Fake Journals.” *TheScientist.com* (May 7): http://www. the-scientist.com/blog/display/55679/.

Gustafson, T. 1975. “The Controversy Over Peer Review.” *Science* 190(December 12):1060–1066. doi:10.1126/science.190.4219.1060.

Hagstrom, O. W. 1965. *The Scientific Community*. New York: Basic Books.

Haug, J. C. 2015. “Peer-Review Fraud—Hacking the Scientific Publication Process.” *New England Journal of Medicine* 373(December 17):2393–2395.

Kronick, D. A. 2004. *“Devant le Deluge” and Other Essays on Early Modern Scientific Communication*. Lanham, MD: Scarecrow.

Laakso, M., P. Welling, H. Bukvova, L. Nyman, B.-C. Björk, and T. Hedlund. 2011. “The Development of Open Access Journal Publishing from 1993 to 2009.” *PLoS ONE* 6(6):e20961. doi:10.1371/journal.pone.0020961.

Larivière, V., S. Haustein, and P. Mongeon. 2015. “The Oligopoly of Academic Publishers in the Digital Era.” *PLoS ONE* 10(6):e0127502.

Latour, B., and S. Woolgar. 1982. “The Cycle of Credibility,” in *Science in Context. Readings in the Sociology of Science*, edited by Barry Barnes and David Edge, 34–43. Cambridge, MA: MIT Press.

McCain, W. K. 2014. “Obliteration by Incorporation,” in *Beyond Bibliometrics*, edited by Blaise Cronin and Cassiry R. Sugimoto, 129–149. Cambridge, MA: MIT Press.

McGuigan, S. G., and D. R. Russel. 2008. “The Business of Academic Publishing: A Strategic Analysis of the Academic Journal Publishing Industry and Its Impact on the Future of Scholarly Publishing.” *Electronic Journal of Academic and Special Librarianship* 9(3): http://southernlibrarianship.icaap.org/content/v09n03/mcguigan_g01.html.

Merton, R. K. 1973. *The Normative Structure of Science*. Chicago: University of Chicago Press.

Merton, R. K. 1988. “The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property.” *Isis* 79:606–623.

Peiffer, J., and J. P. Vittu. 2008. “Les Journaux Savants, Formes de la Communication et Agents de la Construction des Savoirs (17e-18e siècles).” *Dix-Huitème Siècle* n° 40:281–300.

Power, M. 2000. “The Audit Society—Second Thoughts.” *International Journal of Auditing* 4(1):111–119.

Price, D. J. de Solla. 1963. *Little Science, Big Science*. New York: Columbia University Press.

Priem, J., D. Taraborelli, P. Groth, and C. Neylon. 2010. “Altmetrics: A Manifesto.” October 26: http://www.altmetrics.org/manifesto.

Pyenson, L. 2008. “Physical Sense in Relativity: Max Planck Edits the Annalen der Physik, 1906–1918.” *Annalen der Physik* 17(2–3):176–189.

Rennie, D. 2016. “Let’s Make Peer Review Scientific.” *Nature* 535(July 7): 31–33.

Tenopir, C., and D. W. King. 1997. “Trends in Scientific Scholarly Journal Publishing in the United States.” *Journal of Scholarly Publishing* 28(3):135–170.

Thatcher, G. S. 1995. “The Crisis in Scholarly Communication.” *Chronicle of Higher Education* (March 3):B1–B2.

Thompson, J. 2005. *Books in the Digital Age: The Transformation of Academic and Higher Education Publishing in Britain and the United States*. Cambridge: Polity Press.

Van Noorden, R. 2013. “Brazilian Citation Scheme Outed. Thomson Reuters Suspends Journals from Its Rankings for ‘Citation Stacking.’” *Nature* 500(August 29):510–511.

Wade, N. 1975. “Citation Analysis: A New Tool for Science Administrators.” *Science* 188(4187):429–432.

Waters, L. 2004. *Enemies of Promise: Publishing, Perishing, and the Eclipse of Scholarship*. Chicago: Prickly Paradigms Press.

Wilson, T. D. 2002. “The Nonsense of ‘Knowledge Management.’” *Information Research* 8(1): http://InformationR.net/ir/8–1/ paper144.html.

Zuckerman, H., and R. K. Merton. 1971. “Patterns of Evaluation in Science: Institutionalization, Structure, and Functions of the Referee System.” *Minerva* 9:66–100.

3

Playing and Being Played by the Research Impact Game

Michael Power

In 2014, two philosophy departments, at the universities of Birmingham and Keele in the United Kingdom, came equal top in a league table, each with grade point averages of 3.80 (where 4 is a maximum). The Department of Philosophy at Oxford University was ranked tenth with a grade point average (GPA) of 3.40.[1]

Ranking systems are widespread, not least in the field of education (see Kehm, this volume, chapter 6). For example, Espeland and Sauder (2007) examine the rankings of US law schools and their effects on the behavior of key organizational participants, such as deans, who are compelled to pay attention to them despite being doubtful of their worth. Furthermore, while small differences in GPA calculations can amplify differences in rank ordering, these crude snapshots of relative performance provide easy and popular comparability for nonspecialist publics. However, there is something particularly distinctive about the ranking of UK philosophy departments described above: it is based on an evaluation of the *impact* of their research.

By impact in this context, one would ordinarily imagine journal citations and other demonstrable measures of quality within the field of academic philosophy. Such bibliometrics have attracted considerable attention from analysts (e.g., Gingras, 2016). Yet this would be wrong. Impact in this UK setting means the social and economic beneficial impact *outside* academia. In other words, the departments of philosophy at Birmingham, Keele, and elsewhere in the United Kingdom were graded and ranked in terms of the social and beneficial impact of their research. In fact, all subject areas in UK universities were evaluated for this kind of impact as part of a major evaluation of research quality, the Research Excellence Framework (REF2014 hereafter, which is the successor to the Research Assessment Exercises of previous decades). UK universities made 1,911 submissions across all subject areas from 52,061 staff who produced 191,150 research "outputs" of one kind or another. Importantly, for the purposes of this chapter, UK universities submitted 6,975 case studies to demonstrate the impact of their research.[2] As noted above, the Birmingham and Keele departments of philosophy came top of all philosophy departments for their impact.

How is it possible to produce such strange organizational facts as these, and for them not to be regarded as strange? In the next two sections, I provide a brief account of the REF2014 evaluation regime and of the impact case study (ICS) as a new accounting instrument in UK higher education. The argument then focuses on a specific style of evidence gathering for impact adopted by many UK academics—solicited testimony—and uses this feature of REF2014 to suggest that the requirement to demonstrate the impact of research is a *meta-game*, understood as an infrastructure for the production of a certain kind of truth (Foucault, 1980) that also contains the seeds of a kind of "institutional wrongdoing." Indeed, as noted by Wouters (this volume, chapter 4), a meta-game signifies that it has become increasingly difficult to distinguish between gaming the system and the normal functioning of the system (see also Biagioli, 2016). Two important internal features of this meta-game are highlighted. First, while ICSs are qualitative narratives, they are also reductive in nature and create the platform for commensuration via evaluative metrics. Second, the impact meta-game is constituted by a "logic of auditability" with consequences for academic habits and orientation.

REF2014 and Impact

The periodic evaluation of research in the United Kingdom every six to seven years is well established and has been replicated in other countries. Governments demand these evaluations in order to demonstrate that money is being well spent on high-quality research and, in theory at least, to allocate scarce resources to the best universities as judged by the evaluation exercise. In practice, compromises are always required in order to avoid a winner-takes-all concentration of reward in a few universities, leading many to question the purpose of these evaluations.

Different policy fields have had a longstanding interest in the impact of interventions, and "impact" and "impactfulness" have emerged as values in many areas, such as environmental impact assessment. In the UK higher education setting, a decisive catalyst for change was the Warry Report (2006), which recommended that universities measure the impact of their research outside the academy. This ambition finally became a reality for REF2014, and it was decided that twenty percent of total funding for research would be awarded on the basis of such impact (Funding Councils, 2011). Despite initial skepticism and opposition to the crudity of an impact measure for the humanities, UK universities got on with the job of operationalizing the requirement. However, the approach that emerged was very different from the metrics-based regime imagined by the Warry Report. After a period of consultation and experimentation, the regulators of UK universities settled on a case study approach to the demonstration and evaluation of research impact. In effect, while the journal article has long been an "accounting unit" for individual academics (see Wager, this volume, chapter 17; Gingras, this volume, chapter 2), a further new statement was created—the *Impact Case Study* or ICS—to account for the impact of these primary accounting units. And, as noted above, UK universities produced nearly seven thousand of these ICSs.

The Impact Case Study as an Accounting Unit

Academics have a long history of being involved, to a greater or lesser extent, in the world beyond academia. They may be advisers to government and business; they may develop beneficial technologies in medicine and engineering; they may be public commentators on issues of the day and so on. Yet the UK impact regime has changed the status of all these activities, which we might call *engagement*. Simply put, they have come to be regarded as not necessarily impactful in and of themselves. They have been redescribed as "pathways to impact" and therefore as distinct from ultimate impact in the sense of a beneficial change. Accordingly, UK academics have had to ask themselves two questions: "What has changed (outside academia) as a result of my research?" and, crucially, "How can I demonstrate it?"

UK universities needed to provide resources to support the ICS process and to build an infrastructure to cope with this entirely new requirement (Power, 2015). The workload was mitigated to the extent that not all research was required to demonstrate its impact; a norm emerged that roughly one ICS would need to be produced for every ten members of research-active academic staff (this is not a natural ratio and there is no reason to suppose it will not change). Furthermore, in contrast to the Warry Report, the rules published by the regulator were pluralistic about the kinds of impact that academic research might have. Creating a new life-saving drug might be the gold standard, but critical interventions in public policy debates contributing to change would also count as long as they could be proven. Finally, UK universities were highly motivated to comply with the requirement since the financial reward for a 4* ICS—the highest grade—was considerable. Indeed, an official was heard at a conference to confess that the ICSs were "overpriced." Research impact had in effect become big business for UK universities.

Many UK academics, believing or knowing themselves to be impactful, embraced the impact agenda and got on with the work of producing their case studies. They were advised and quality controlled by committees that were themselves learning the impact game. ICSs would be evaluated for their "reach" and "significance" and there was a prescribed template design to shape and limit the form, content, and length of an ICS. Although ICSs might essentially be narrative in form (i.e., be "stories of impact"), they would have a prescribed structure with maximum word counts for each section. There were also more specific rules—not least about the accounting time window within which impact might be measured—but also quite a lot of pluralism about the kinds of possible impact to consider and also about the forms of evidence that might be used to support claims of impact (Funding Councils, 2011). A particularly significant form of evidence used by scholars at the London School of Economics (LSE), including myself, was "solicited testimony." Reflecting on this evidence form provides some insight into the workings of the research impact regime.

Evidence of Impact and Solicited Testimony

In the practical field of evaluation, it is accepted that a way to find out if an intervention has had an impact is to ask those people or groups who one would naturally expect to have been "impacted." In other words, their testimony is solicited. Although this method of data collection has its own epistemological shortcomings, and care must be taken not to lead the respondents, for very pragmatic reasons, it was an attractive form of evidence for social scientists. Many very smart people at LSE encountered problems writing their ICSs largely because of the difficulties of causal attribution coupled with normal scientific caution and modesty. And even where their external impact might be self-evident to themselves, it would not be to others. Because the effects of research, if any, dissipate into the wider social and institutional environment of universities, evidential traces of impact lie outside the organization and are costly to collect. At LSE, the ICS proved to be a distinctive kind of genre that had to be crafted, and the process gave rise to many unexpected difficulties. In such a setting, and with a race against time to prepare the ICSs, the epistemic weaknesses of solicited testimony were outweighed by its pragmatic, low-cost features.

Terms like “solicited testimony” imply seriousness and scientificity. But the impact game was new, and many academics, in the search for corroborating evidence, rang or emailed those who may have been influenced by their research. Conversations might have gone like this: “Look, you know that piece of work I did for you, based on my research, which you thought was important? Would you mind just putting that in a letter to me or in an email please?” In this way, a trace of external impact was created that could be collected as evidence. Importantly, it was not sitting out there waiting to be found. It was actively constructed by the researcher.

Meta-Gaming

What is going on here? Were academics like me gaming the system by seeking solicited testimony in this way? Was this a cheap and quick way to build a credible ICS, particularly as the narrative of impact had to be constructed *ex post* and in a hurry? This is undoubtedly partly the case; at LSE, concerns were voiced about the overuse of this evidence type in many ICSs. These concerns were also shared by evaluators. One report on the REF2014 process and the evaluation of ICSs notes that they found it “very hard to assess the significance of an impact where evidence was nuanced and in the form of corroborating testimonials” (Rand Europe, 2015). However, much more is at stake here than gaming by individual academics.

The impact agenda in the United Kingdom represents a rebalancing of two logics or values that have always existed in tension with each other in academic life, namely, the logic of autonomous curiosity and the logic of use-value and economic benefit. Furthermore, different universities and different subject areas will combine these logics in different ways. Fields such as social policy, where action research is common, embraced the turn toward impact. Strangely perhaps, in business schools, where subdisciplines have been on a path to greater academic respectability, the impact agenda created more challenges than might be expected. And for fields such as history and philosophy, the agenda was entirely new and disruptive.

However, while the rise of impact accounting in UK universities exhibits this diversity of reactions, the example of solicited testimony as an evidence form suggests that something systematic is at stake. Following REF2014, new academic habits are visible, supported and routinized by new infrastructures and databases for the collation and analysis of impact. Universities are creating dedicated roles, such as "impact officers," and providing support to academics to help them maximize their impact (LSE Public Policy Group, 2012). Research funding bodies are awarding prizes for impact. In short, an entire *impact apparatus* is being created. UK funding bodies also often require a statement of *expected impact* or something similar in grant applications. In effect, this means that applicants must turn the causal pathway model on its head; the economic stakes of being successful in raising funding implicitly requires them to ensure that they are having or will have an impact *before* they do the related research. So the grant application process reveals a variant of Goodhart's law, whereby an *ex post* outcome measure of impact "flips" to become an *ex ante* target (see also Griesemer, this volume, chapter 5, for more on Goodhart's law). Rather than impact being a measured outcome of research, as it was for REF2014, it is now research, or a certain style of research, that is becoming the product of an apparatus that *targets* impact (Power, 2015).

Formally not all UK academics are required to prepare ICSs and be impactful. Yet impact is now one of the formal criteria in promotion guidance. Impact has therefore become an established norm of evaluation regardless of nuances of scope. Consequently, as the example of solicited testimony suggests, UK researchers are learning how to engage with possible users of their research, constructing them as good impactees who are actively cultivated and internally represented for the purpose of writing future ICSs.

From this point of view, the use of solicited testimony is not simply the gaming of the impact accounting system by individual academics. Rather, it reveals how the system actually works. The impact accounting regime may indeed be subject to gaming of its rules, but those rules in aggregate, embedded in apparatuses, constitute a *meta-game*, namely rules for the production of a certain kind of truth in Foucault's (1980) sense, meaning systematized ways of governing what it is possible for individuals to say, and what actions and performances are legitimate. While it is easy to hold specific humans to account and make them visible *within* an accounting system, it is much harder to hold an accounting system to account and to provide a critical account of its operating logic. Below I argue that this operating logic in the case of research impact in the United Kingdom has little to do with social and economic benefit—the originating policy values—but is motivated by a distinctive cultural commitment to evidence gathering and its associated disciplinary power. I call this the "logic of auditability" (Power, 2019).

Qualitative Commensuration

While the ambition of the original Warry Report was for the measurement of impact, and this is an ambition which has not disappeared, the ICS regime that took shape in the United Kingdom did not take this direction. While metrics (e.g., citations in public documents) supporting impact may be used as forms of evidence, this is within the context of an overall case study narrative. However, if metrics are not a required input of the impact accounting system, they are a consequence of it. The grade point average of 3.8 scored by the departments of philosophy in the universities of Birmingham and Keele noted above are composites of scores given to individual ICSs by a panel of evaluators. So at this point of evaluation, very different ICSs narrating very different kinds of impact become commensurable, are made capable of being compared, and, importantly, ranked (Espeland and Stevens, 1998).

The process of commensuration, the point at which qualities are made into quantities, is a continuing source of interest to scholars and to the emerging "valuation studies" agenda.[3] However, the case of research impact in the United Kingdom suggests that there may not always be a singular ontological jump from quality to quantity. The ICS accounting regime provides an example of how qualitative characteristics become subject to quasi-commensuration in templates and requirements for narrative precision. This is not a direct form of quantification, but rather its conditions of possibility—in this case, it enables grading by an evaluator. In other words, the point of metricization and quantification has its conditions of possibility in the construction of qualitative or hybrid narrative forms that integrate with, and support, quantification. So the ground for commensuration has already been prepared in their construction. The final stage of quantitative commensuration can then be performed by evaluators who make judgments about the quality of each ICS and express them in metrics.

Impact and the Logic of Auditability

The research impact requirements in the United Kingdom emerged from policy ambitions to reconnect universities to the UK economy and to make this demonstrable. In particular, the ICS infrastructure requires the traces of impact to be constructed and reported in templates that enable them to be evaluated and audited. Such a model is familiar to accountants—it's the way accounts are produced. Yet, this creation of traces, how we do it and for what activities, whether they're citations or other fact-bearing objects, reflects a cultural logic of a distinctive kind, a logic of auditability (Power, 2019). Fundamentally, this is a *logic of trace production* rather than measurement specifically. For example, solicited testimony discussed earlier is one form of systematic, nonquantitative trace construction that is interiorized, embodied in ICSs, and then evaluated. Guided by the logic of auditability, the trace does not simply record the impact; it defines and constitutes what impact is.

The logic of auditability as a cultural form of trace production matters when considering the evaluators of the 6,975 impact case studies noted above. It is clear that that they found the evaluation process very difficult, despite the discipline of the ICS production process. Evaluators reported that they had very little time or resources to drill down to check (i.e., audit) the underlying sources or traces referenced in the ICSs, and they drew attention to the poor quality of links to underlying evidence (Rand Europe, 2015). Their operational challenges reveal the face of the real "audit society" (Power, 1997). It is far from being a fully transparent and "auditable" society, or even one that's full of very confident inspectors and watchdogs who perform checks. Rather, the logic of auditability is essentially *productionist*; it names a cultural compulsion for organizations to construct and collect traces of activities, and to fabricate audit trails that link accounts of performance, such as ICSs, to an underlying evidence base. That evidence must be shaped, selected, and constructed so that traces of impact are auditable in principle, even if this is not possible in actuality.

The logic of auditability is a cultural value that passes for common sense and is reinforced by the construction of infrastructures for trace production. In this respect, the impact regime is a meta-game for the production of a distinctive kind of facticity. It is a meta-game because it embodies a potential for misconduct that has little to do with gaming metrics at the individual level. Rather, a kind of misconduct is embodied in the infrastructure itself whose deep logic is that of auditability and trace production, whether the traces are citations within academia (Biagioli, 2016) or traces of impact in the "outside" world. The logic of auditability does not itself value users or consumers or indeed economic value as such—the value is in the trace (Power, 2019). To understand the power of this logic is to understand why there seems to be so much investment in evaluation systems that are very costly and seem to produce little economic benefit.

Conclusions: From Traces to Metrics

The chapter began with the example of UK philosophy departments playing the game of research impact. Of course we know that philosophy has an impact; the influence of Plato and Aristotle on Western culture and cognitive assumptions is well known. However, had Plato and Aristotle lived today, I think they would have been hard pressed to construct an ICS and to collect traces of the kind required by the REF2014 exercise. Ironically, this is because their impact has been so great and so diffuse that even the most helpful impact support officer would struggle to support them in producing the evidence. This is another paradoxical feature of the impact regime in the UK: the more impact you actually have, the harder it is to account for it.

I have suggested that the regime for accounting for research impact in the United Kingdom is a kind of meta-game in the form of elaborate rules for the production of a new kind of truth about academic research. Apparent individual gaming of this regime, such as the use of solicited testimony discussed above, is in fact a feature of the way the regime works. Following Foucault (1980), I also allude to a growing "impact infrastructure" that regulates the production of this truth in the form of acceptable "accounting" statements about impact. This impact infrastructure is in turn itself permeated by a systemic logic of auditability that demands the production of traces (not just numbers), which, when gathered into the ICS template, can be evaluated and then metricized. As regimes of impact truth production, these infrastructures have also created new forms of deviance. Strangely, this deviance does not always take the traditional form of academic fraud or gaming the system. Rather it also involves carrying on as before without regard for the cultural imperative to produce precise traces of activity that can "travel." From this point of view, being an intellectual and writing long books are forms of deviance relative to, and brought into existence by, the research impact accounting system in the United Kingdom. But in a system whose logic would define Plato and Aristotle as deviant, we should probably ask where the ultimate misconduct really lies.

Notes

1. https://www.timeshighereducation.com/sites/default/files/Attachments/2014/12/17/k/a/s/over-14-01.pdf. Accessed on May 9, 2016. A grade point average is simply what it says (i.e., an overall average of grades given on individual units of evaluation, usually calibrated from 1 to 4). The impact of all subject areas

was aggregated to give a score for impact at the individual institutional level. The UK Institute of Cancer Research was top in the rankings overall.

2. See http://www.ref.ac.uk/media/ref/content/pub/REF%20Brief%20Guide%202014.pdf. Accessed on May 9, 2016.

3. For the recently created online journal *Valuation Studies*, see http://valuationstudies.liu.se/.

References

Biagioli, Mario. 2016. "Watch Out for Cheats in the Citation Game." *Nature* 535:201.

Espeland, Wendy, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113:1–40.

Espeland, Wendy, and Mitchell Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24:313–343.

Foucault, Michel. 1980. "Truth and Power." In Colin Gordon (ed.). *Michel Foucault Power/Knowledge: Selected Interviews and Other Writings*. Brighton, UK: Harvester Press.

Funding Councils. 2011. *Decisions on Assessing Research Impact*. Bristol, UK: HEFCE.

Gingras, Yves. 2016. *Bibliometrics and Research Evaluation: Uses and Abuses*. Cambridge, MA: MIT Press.

LSE Public Policy Group. 2012. *Maximizing the Impacts of Your Research: A Handbook for Social Scientists*. London: London School of Economics Public Policy Group.

Power, Michael. 1997. *The Audit Society*. Oxford: Oxford University Press.

Power, Michael. 2015. "How Accounting Begins: Object Formation and the Accretion of Infrastructure." *Accounting, Organizations and Society* 47:43–55.

Power, Michael. 2019. "Modelling the Microfoundations of the Audit Society: Organizations and the Logic of the Audit Trail." Academy of Management Review. https://journals.aom.org/doi/pdf/10.5465/amr.2017.0212.

Rand Europe. 2015. *Assessing Impact Submissions for REF 2014: An Evaluation*. Cambridge: Rand Corporation.

Warry Report. 2006. *Increasing the Economic Impact of Research Councils: Advice to the Director General of Science and Innovation from the Research Council Economic Impact Group*. London: Department for Business, Innovation and Skills.

4

The Mismeasurement of Quality and Impact

Paul Wouters

In this chapter, I would like to discuss the implication of the problematic relationship between "gaming the system" and "properly operating in the system" for research evaluation. As Biagioli and Lippman (Introduction, this volume) argue, the current audit culture in academia has created new dimensions of academic misconduct. The performance indicators that are supposed to objectively measure the value of scientific production have become the objects of new forms of manipulation. Because research evaluations have become reliant on proxy indicators that measure only indirectly what they are supposed to represent (either quality, impact, or societal relevance), it is increasingly hard to see the difference between indicator manipulation and authentic high-value research. According to Biagioli and Lippman (Introduction, this volume), "the traditional locus of evaluation—the publication's claims—has become technically irrelevant to metrics regimes based on impact factors." Reading has stopped being necessary in institutional evaluations, they argue, while the role of metadata has shifted from descriptive to evaluative. While I think the practice of evaluation is still quite messy (in many assessments, reading might still happen), and the role of indicators is not so straightforward (the Journal Impact Factor, though the most popular indicator, is not the only game in town and neither representative for all), I do agree that Biagioli and Lippman very perceptively sketch a portrait of the dominant tendencies in the academic audit culture and of the market enabled by the infrastructure of citation (Wouters, 2014). Indeed, if these trends remain dominant, high performance scores will be *made* identical to high-value work. Those aspects of academia that are not presentable in indicators may then no longer be supported (by researchers themselves as well as by their employers) and may have to migrate to the area of works of love (amateurs)—or become extinct. Scientific research may in such a scenario still lead to exciting innovations, but the notion of knowledge itself as an important aspect of human culture would no longer be an important *leitmotiv* of universities. Scientific research will then have been instrumentalized in its totality.

If we wish to somehow preserve, or restore in new forms, this aspect of academic life, it seems useful to imagine alternative scenarios. This should also involve alternative forms of research assessments and performance evaluation of individual researchers, research groups, and institutions. Imagining these will require a bolder redesigning of current evaluation protocols at universities than is currently the case. In particular, I will argue, evaluation experts and scientometrians need to go beyond the popular concept of "informed peer review" (Moed, 2005). It is not sufficient to claim that peer review and indicators need to be combined in intelligent ways because the very basis of what counts as an intelligent combination is at stake. What is needed is a more radical *recontextualization* of indicators as well as qualitative evidence in assessments.

Before discussing a possible alternative to the market-oriented, indicator-driven forms of evaluation, it makes sense to briefly outline how research assessments have become influential in shaping academic life.

First of all, the conduct of research has become so strategic that it is vital for researchers to be visible at both the national and international level. Second, partly as a result of the success of scientific and technological research in many fields, and partly as the consequence of independent policy developments, the traditional academic autonomy no longer exists. Apart from some exceptions, the scholarly community on which Robert Merton built his sociology of scientific norms (Merton, 1973) has become pervaded by extra-academic interests and communications (Leydesdorff, 2000; Shapin, 2008). Third, research groups and individual researchers in the public research system in all disciplines are subjected to recurring institutional assessments in which performance must be made visible in the terms of that specific institution. Usually, scientific quality (or excellence) and societal impact are the main pillars. Other criteria such as the quality of teaching and PhD training, viability and feasibility of the research plans, and, last but not least, earning power are drawn in as well. The key issue is that the results of these evaluations produce the symbolic capital with which the researchers can—or cannot—participate in the next cycle of research.

Two different forms of research evaluation are usually distinguished in opposition to each other in the interdisciplinary debate about the best way to assess academic research and universities. The first is the qualitative judgment called peer review, based on the assessment by scientific experts usually working in the same or a related field as the research group. The second is assessment by indicators often based on simple or complex forms of citation analysis in which high numbers of citations are seen as proxy for influence or quality. Academia has a lively debate about the weight each form of evaluation should have and about the extent to which it is possible to combine one with the other (Hicks et al., 2015; Wilsdon et al., 2015).

In this debate, it is often overlooked that the two forms of assessment are intrinsically and intensely linked to each other. This was not yet the case when the Science Citation Index was invented by Eugene Garfield (Garfield, 1955; Wouters, 1999, 2014; Csiszar, this volume, chapter 1). But as a result of the rise of the use of citation-based performance indicators since the early 1980s, first in national science policies and later in the management of universities and research institutes and in global university rankings, both methods are no longer purely quantitative versus qualitative but have intertwined and interpenetrated each other. This is relevant to the debate about research integrity and norms for proper scientific behavior because, as we will see, comparable forms of mixing are making the identification of improper behavior less evident. So it is worthwhile to spell out the connections and mutual pollution between peer review and assessment by indicators in more detail.

Citing relationships are, in the end, based on the decisions by authors of scholarly papers to include formal references in their bibliography to scientific work that they deem relevant (implicit references do not end up in citation indexes, although they may be very important intellectually). These citing relationships are usually concentrated within the same research area combined with additional interdisciplinary connections to other literature. The decision to cite a specific reference, and not an alternative piece of work, is shaped by a complex mixture of intellectual and strategic motives. The guidelines of scholarly journals or book publishers regarding number and type of references form the template for these decisions, but there is still a lot of leeway for the authors to express their preferences. It is difficult, if not impossible, to clearly separate intellectual and strategic motives. In addition, it should be stressed that both motives are of a social nature.

In addition to this basic connection at the individual level, there is also a group connection between peer review and citation-based indicators: they draw upon the same scientific or scholarly community. This may vary by type of document and by discipline, but, very often, the researchers in the citation network are also involved in the regular peer-review work of journal and book publications. An interesting question, not yet frequently studied, is to what extent these communities are also the basis for post-publication peer review such as the national research assessment exercises. The latter forms of peer review may draw upon more interdisciplinary networks and hence be more removed from the core peer networks in particular research areas.

A third connection between citation and peer-based evaluation is the reflexive loop that has been created by the emergence of citation-based performance indicators. Because researchers have become aware, on a large scale, that their bibliographies may influence the careers of the researchers they cite, their "citing behavior" will be affected by this knowledge. The strategic motives may therefore have become more important because the competition among researchers has been extended to the domain of metadata such as numbers of citation. Ethnographic research of publication and evaluation practices has shown that researchers tend to reason quite strategically about both their publication outlets and their bibliographies, although it is also clear that this varies strongly by field and type of research (Rushforth and de Rijcke, 2015). This does not mean, *inter alia*, that bibliographies have become completely dishonest and unreliable as a source for intellectual queries, but it has certainly made the sociological interpretation of citation frequencies and networks more complex. It is an example of the basic reflexive nature of communication behavior and networks (Leydesdorff, 1995).

This third connection may have important consequences for the conduct of post-publication evaluation of research performance. Most formal evaluation protocols do not include instructions to use citation indicators and some even discourage the use of indicators such as the Journal Impact Factor or the Hirsch Index. However, this does not mean that these indicators do not play a role in, for example, the preparation of a committee session by individual evaluators. Because the citation indicators have basically become an easily available resource and can indeed be seen as a citation infrastructure (Wouters, 2014), consulting Web of Science–, Scopus-, or Google Scholar–based citation scores may be matter of course without much conscious deliberation. They have even led to "folk citation theories" upon which researchers draw in their interpretation of these data (Aksnes and Rip, 2009; Rushforth and de Rijcke, 2015; Wilsdon et al., 2015). For example, biomedical researchers are usually quite aware of the technical limitations of the impact factor and use these as components of a field-specific interpretation of the journals. High impact factors may then be interpreted as indicating journals with a large number of submissions and a high rejection rate. Hence, publishing in these journals is a sign of success in the fierce competition for recognition. Rushforth and de Rijcke (2015) show how the impact factor is used in comparable ways as a judgment device that is already deeply engrained in collaboration and publication strategies. This does indeed indicate that the impact factor, as Biagioli and Lippman (Introduction, this volume) argue, functions as a measure of value in a market, in other words as a currency.

In the recent Higher Education Funding Council for England report *The Metric Tide* (Wilsdon et al., 2015), a number of norms for proper evaluation have been proposed. The unifying concept here is "responsible research metrics," which makes an important reference to the European policies on "responsible research and innovation." Responsible research metrics should be seen as those practices in using quantitative performance indicators that are attuned to five core principles:

- **Robustness**: basing metrics on the best possible data in terms of accuracy and scope
- **Humility**: recognizing that quantitative evaluation should support—but not supplant—qualitative, expert assessment
- **Transparency**: keeping data collection and analytical processes open and transparent, so that those being evaluated can test and verify the results
- **Diversity**: accounting for variation by field, using a variety of indicators to reflect and support a plurality of research and researcher career paths
- **Reflexivity**: recognizing the potential and systemic effects of indicators and updating them in response

These principles have been formulated on the basis of the recognition of the complex interplay between peer judgment and citation-based indicators. If national and institutional research assessments and the building of the databases used in them would consistently adhere to these principles, it would surely represent important progress in evaluation practices. But how should we interpret the second principle, "humility"? It can easily be read as the incorporation of quantitative indicators within a framework that is dominated by peer-review and expert judgment. But is this really a form of humility?

The recent Leiden Manifesto formulated a comparable norm for the use of quantitative performance indicators (Hicks et al., 2015): "Quantitative evaluation should support expert assessment." The manifesto is a warning against exaggerated forms of performance-based indicators and pleads for a judicial combination of quantitative and qualitative evidence in research evaluation. This basically builds forth on the concept of "informed peer review," which also aims to combine citation analysis with peer review, both in the context of peer review itself (where indicators would simply be a form of evidence next to other forms of qualitative or quantitative evidence) and as a check on the integrity of the peer-review process itself (Moed, 2005; Butler, 2007).

Informed peer review can be seen as an attempt at triangulation: if we can collect more evidence and this evidence points in the same direction, surely we have a more robust foundation for our conclusions? However, this is based on the assumption that the two forms of evidence are independent of each other. As we have seen, this is only partially the case. In addition, a practical problem arises when citation analysis and peer judgment are in conflict with each other about a particular research performance. On what basis should the evidence be weighted? The idea that expert judgment is always better ignores the gatekeeper role of scientific reviewers and is naïve with respect to the strategic motives operating in the process of peer review. Conflicting outcomes of peer review and citation analysis may for example be a signal that the reputational mechanisms in the academic system are not keeping up with novel developments in research. Peer review may also play in the hands of "old boys" networks and discriminate against women and ethnic and intellectual minorities in science. Relying mainly on peer review may delay interdisciplinary innovation because this often entails not only a reconfiguration of substantive or methodological research areas, but may also mean a redefinition of the very criteria of what counts as high quality in research. In other words, peer review in a particular discipline may simply filter out radical innovation because it is not recognized as high quality.

So we cannot always rely on peer review as the ultimate arbiter. But the same holds for quantitative performance indicators. Neither the number of publications nor their number of citations, normalized for field, document type, and age or not, can simply be interpreted as proxies for quality or impact. Researchers with exceptional publication numbers may be opening up an exciting new field or they may be very good in gaming the performance system. A high number of citations may indicate great research with a huge influence or they may come from humdrum me-too research or even citation cartels. And a low number of citations may result both from less interesting research and from path-breaking studies that are not yet recognized. So we need some form of judgment to assess the value of publication or citation performance criteria (the same holds for indicators of earning power). But this brings us back to the experts involved in the peer-review system. We seem to be caught between a rock and a hard place.

And yet, this does not hinder assessments to take place. In fact, the international research system is a buzzing evaluation machine (Dahler-Larsen, 2012). The construction of peer review and metrics-based assessment as two opposites, common in generalized debates about research evaluation and in most research policy discussions, is a false one in the sense that it is not what happens in the varied practices of research evaluation. Rather than the dichotomy of qualitative versus quantitative, or peer review versus measurement, we should focus on the context of evaluation. The tendency to speak of research evaluation as such tends to ignore the wide variety of practices that are bundled in this container concept. The weight of impact indicators, for example, varies strongly between the assessment of the results of a PhD student and the ranking of a university in comparison to its international peers. The tendency to speak of research evaluation as a somehow integrated institution is a form of purification that tends to make invisible precisely that which should be foregrounded in an alternative discourse. In this light, the proposals to formulate principles of responsible metrics and responsible evaluation are only a first step. These could still easily be incorporated in the framework of a market-oriented audit culture. To enable a true alternative to the dominant trends in academic research evaluation, we need to complement the concept of responsible metrics with the recognition that valuing is principally an act of judgment in context. Decontextualized information, whether peer based or indicator based, needs to be put back into context if we wish to create a strong barrier in assessment practices to academic misconduct in all its novel forms.

Technically, my proposal is to replace the notion of "informed peer review" as the supposedly most nuanced approach in research assessments by "contextualized judgment," which 1) puts context central and 2) does not create a false dichotomy between peer review and indicator-based assessment. It takes into account the flexible, and often quite ingenious, ways in which researchers attach meaning to constructs like peer opinion and citation indicators (Rushforth and de Rijcke, 2015).

Politically, following this approach would put two questions central in the construction of new evaluation protocols and procedures:

1. How will this evaluation design influence the creative process of knowledge creation?
2. Who is in control of the agenda setting and the research process?

The first question is about coping with matters of perverse effects, the alignment of the criteria in the evaluation and the mission of the specific research group or program, and effects on the texture of power in the field. The second question addresses the way quality is managed in the field and its connections with stakeholders, nonacademic partners, and society at large. Both questions point to the political aspects of norms for evaluation. This is as it should be since matters of evaluation are deeply political matters and deal with the question of how we wish to live and what kind of society we are creating (Mol, 2002; Thurtle and Mitchell, 2002).

Notes

Acknowledgement: I would like to thank Thomas Franssen, Sarah de Rijcke, and the referees and editors for their comments on an earlier version. This work was partly funded by the Norwegian Research Council through the Center for Research Quality and Policy Impact Studies (R-QUEST) (https://www.r-quest.no/) and by the Riksbankens Jubileumsfond (Sweden) through the KNOWSCIENCE project (https://www.fek.lu.se/en/research/research-groups/knowscience).

References

Aksnes, D. W., and A. Rip. 2009. "Researchers' Perceptions of Citations." *Research Policy* 38(6):895–905. http://doi.org/10.1016/j.respol.2009.02.001.

Butler, L. 2007. Assessing University Research: A Plea for a Balanced Approach." *Science and Public Policy* 34(8):565–574. http://doi.org/10.3152/030234207X254404.

Dahler-Larsen, P. 2012. *The Evaluation Society*. Stanford, CA: Stanford University Press. Retrieved from http://www.amazon.com/The-Evaluation-Society-Peter-Dahler-Larsen/dp/080477692X.

Garfield, E. 1955. "Citation Indexes for Science Through Association of Ideas." *Science* 122(3159):108–111.

Hicks, D., P. Wouters, L. Waltman, S. de Rijcke, and I. Rafols. 2015. "The Leiden Manifesto for Research Metrics. *Nature* 520:429–431. http://doi.org/10.1038/520429a.

Leydesdorff, L. 1995. *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications*. Leiden, The Netherlands: DSWO Press.

Leydesdorff, L. 2000. "A Triple Helix of University-Industry-Government Relations." *The Journal of Science and Health Policy* 1:43–48.

Merton, R. K. 1973. *The Normative Structure of Science*. Chicago: University of Chicago Press.

Moed, H. F. 2005. *Citation Analysis in Research Evaluation* (Vol. 9). Dordrecht, The Netherlands: Springer.

Mol, A. 2002. *The Body Multiple: Ontology in Medical Practice*. Durham, NC: Duke University Press.

Rushforth, A., and S. de Rijcke. 2015. "Accounting for Impact? The Journal Impact Factor and the Making of Biomedical Research in the Netherlands." *Minerva* 53(2):117–139. http://doi.org/10.1007/s11024-015-9274-5.

Shapin, S. 2008. *The Scientific Life: A Moral History of a Late Modern Vocation*. Chicago: University of Chicago Press.

Thurtle, P., and R. Mitchell. 2002. *Semiotic Flesh: Information and the Human Body*. Seattle: University of Washington Press.

Wilsdon, J., L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, R. Jones, R. Kain, S. Kerridge, M. Thelwall, J. Tinkler, I. Viney, P. Wouters, J. Hill, and B. Johnson. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. http://doi.org/10.13140/RG.2.1.4929.1363.

Wouters, P. 1999. *The Citation Culture*. Amsterdam: University of Amsterdam. Retrieved from http://garfield.library.upenn.edu/wouters/wouters.pdf.

Wouters, P. 2014. "The Citation: From Culture to Infrastructure." In B. Cronin and C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Performance* (pp. 47–66). Cambridge, MA: MIT Press. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10650330.

5

Taking Goodhart's Law Meta: Gaming, Meta-Gaming, and Hacking Academic Performance Metrics

James Griesemer

This chapter offers two thoughts to the meta-conversation about academic performance metrics and misconduct. One is that Goodhart's law (1975) concerns more than simply the idea of individual responsiveness to pressures from societal policies, for example, central bank monetary policies employ economic performance measures as standards of regulation and control in banking. The other concerns how we might exploit what *more* there is to Goodhart's law to probe the character of "mis"-conduct, as individuals and organizations adapt to, and comply with, academic performance metrics institutionalized as standards. Contrast this with "bad" conduct, as individuals and organizations cynically attempt to "game" or "exploit" the system to achieve a better evaluation *than their performance warrants*. Along with other chapters in this volume (Csiszar, this volume, chapter 1; Power, this volume, chapter 3), I suggest Goodhart's law describes conditions that not only undermine the representational success in modeling causal order in human social systems, but also the operation of the law *inverts* the causal order. Conversions of metrics into standards not only *invite* "gaming the system," they also practically construct "gaming" as the new form of practice, rendering the original product or practice to be measured as a "side effect" in the new causal order. Or, as Wouters (this volume, chapter 4) urges, we must distinguish gaming the system from properly functioning in an inverted system. It is thus problematic to moralize and shame so-called "predatory" practices as if it were clear what constitutes ethical, nonpredatory practice in social worlds where Goodhart's law operates.

Goodhart's lesson was that such measures are self-defeating because they invite "mis"-conduct. If people respond to standards as intended, the measure ceases to represent and record the primary target performance and comes to measure only compliance or conformity to the standard. The critique cuts deeper. As Lucas's (1976) critique of macro-econometric models showed, such measures are self-defeating because the underlying causal structures of individual and organizational social behavior change when people and organizations respond to policies based on the models, so the policy causes the model to cease to represent the very thing the measure was designed to measure as it changes the system's causal structure. Goodhart's law and Lucas's critique apply to social policies built upon quantitative metrics taken as standards of quality and performance evaluation, for example, academic achievement. So the causal influence is reciprocal. This creates conditions for an arms race. The only escape from arms races is to realize their futility. Either policy-makers must stop making arms-race–producing policies, or the governed must revolt against intolerable institutions.

The second thought is that, philosophically, asking a different *question* may lead to more valuable insights than seeking *answers* to the question originally posed. Questions like "what is the nature of misconduct in response to academic metrics?" are of this sort because *posing* this question itself reinforces or facilitates circumstances in which Goodhart's law will apply. Framing the question of academic misconduct as one of gaming a system of metrics is to accept metrics as standards, if only for the sake of argument. Asking draws attention to the problem while entrenching presuppositions of the question. It is a prime mode of escalation in a metrics arms race between standards imposers and gamers. Csiszar (this volume, chapter 1) calls this an "inverse" form of Goodhart's law: "Only when a measure becomes a target is it widely taken up as a measure worth using." Practices and policies that use metrics *as* standards turn work performance, including scholarship, into a game in which the goal is to exceed the standard rather than perform the work that was to be measured. In other words, compliance with the standard becomes the goal rather than a side effect of the performance, and, in turn, performance of the work becomes a side effect of a policy imposing a standard. As Chamayou (2009) disturbingly but eloquently argues: the goal is to write articles, not do research. Other institutions are changing in parallel ways, so this is a widespread change across societies. Managing health *risk* has become the goal (and meaning) of health while curing illness has become the side effect of pharmaceutical use, rather than the other way around, as Dumit (2012) has shown.

Alternative questions are: If academic performance metrics are a game, can we hack them? If metrics-based standards and gaming the system create an arms race to nowhere (or to the decline and fall of the research system), what would it take to get rid of them? If metrics-based academic performance standards can be hacked, then perhaps hacking would reveal their futility as unsuited to serve as *standards* to begin with. Perhaps hacking would redirect the question of "mis"-conduct back onto social policies of performance evaluation and quality judgment and away from charges of manipulated metric outcomes and undeserved gain. My aim is to propose interventions leading to better questions, whether or not they suggest good answers to old questions.

Performance and publication metrics are of epistemological interest concerning how scientific practices become transformed through institutional change, not only of ethical concern about the "conduct" of scientists and their publishers. Studies of gaming metrics may provide insight to philosophy of science as well as to research ethics.

Power's chapter and his book, *The Audit Society*, urge us to think about the kinds of problems surveyed in this volume from the point of view of narrative impact stories in addition to quantitative metrics (Power, 1997). The relation between individual narratives and aggregate impacts is a productive way to articulate where and how the "audit society" may have gone off the rails into the kind of self-defeating process Goodhart warned of. Turning a creative individual enterprise into the bureaucratic one of pursuing metrics as the *meaning* as well as the measure of productivity and success, of pursuing the "CV for its own sake," as Biagioli remarked in our workshop, is the moment of translation of performance in an audit society into a potential for "mis"-conduct, that is, conduct that emerges as unethical or unbecoming in the system *as it used to be* but which no longer functions that way. Institutionalized "mis"-conduct then becomes "bad" conduct when the ethics and optics of performance are *judged* against the old system but *measured* in the transformed system. Ethical evaluation lags metrical assessment, and "mis"-conduct that used to signal unethical "bad" conduct should now merely point to this misalignment, not express condemnation. The ethical judgment becomes misplaced and unjustified because the research system is no longer correctly understood epistemically.

Fundamentally, anyone interested in scientific practice must take into account how the research experience is now shaped by academic metrics. The "metric tide" (Wilsdon et al., 2015) demands that philosophy of science consider that science is now conducted in a regime of metrics, not only for research performance evaluation, but also for judgments and decisions affecting workflows in the research process itself: what problems and projects to pursue, what grants to seek, what personnel to hire, and what schools to attend, not merely which journals to publish in.

The comprehensive discussions and critiques in *The Metric Tide* report (Wilsdon et al., 2015) and *The Leiden Manifesto* (Hicks et al., 2015) indicate that there is a kind of balancing act going on in the metrics debate about what the future holds. The war on metrics seems to be over: metrics will not go away, even if which metrics are today's favorites will face a constant churn. We can describe the central tension in various ways, but they boil down to the idea that expert judgment should play a central role in evaluating research content and that quantitative data should play a role in measuring research productivity. Judgment cannot be automated, yet productivity can in some ways be measured.

The Leiden Manifesto says the problem is that performance evaluation is now *led* by the data *rather than* by judgment. If balance between judgment and data is the goal, then the question to ask is how to rebalance, reenergize, and reimagine a role for judgment in the face of the data-driven metrics onslaught. The metric tide can only rise because data now comprises a gravitational force tugging on the digital ocean; the energy of big data makes for ever larger waves. Solving the rebalancing problem will not be achieved by trying to turn back the digital ocean.

To explore how to rebalance judgment in evaluation of research, consider a thought experiment to probe the alleged greater "objectivity" and "reliability" of measurement with data and "subjectivity" of judgment. The experiment is to "go meta" in a *strategic* gaming response to metrics-based evaluation. I take inspiration from Daston and Galison (1992, 2007; Galison, 1998): the image of objectivity is historical, not static. They historicize the concept of objectivity by showing how it flip-flops between mechanical and judgmental zeitgeists.

Thought experiment isn't enough to advance our understanding of the problem sufficiently to pose the right questions, however. I propose actually *doing* the experiment, which I will call "hacking the metrics." If we think of hacking as a form of experimentation and experimenting as a form of research, the idea is that hacking can be both a means of intervening in the metric tide and a legitimate mode of research about it.

Experimentation of this kind should challenge ethical intuition about conduct and at the same time *transparently* undermine metrics for experimental purposes while redirecting questions about performance and rebalancing evaluation toward judgments of the work.

One way to make metrics hacking into a research enterprise is to exploit causal aspects of Goodhart's law. The idea is to reframe the question of gaming as a problem of causality rather than representation, one that can be tested by experimental intervention in ways that might actually change the use of metrics as standards by disrupting the policy debate. If metrics are subjected to hacking, then perhaps the end game of finding an unhackable metric will come to seem a hopeless task to policy makers bent on automating judgment or replacing trust with "objective" quantitative measures—a fool's errand as ridiculous as an uncrackable cryptographic scheme or a winnable nuclear war.

The solution *du jour* to the problem of gaming performance metrics like citation counts, h-indexes, or impact factors is to multiply metrics and form a "basket" of them, each metric serving a particular component job well rather than hoping an all-purpose metric will emerge from the arms race. I am on board with the "alt-metrics" or basket approach as methodological antidote to the idea that some single, best all-purpose metric will be found to replace citation count or impact factor, but I am skeptical that it is going to slacken the metric tide because we are already in a positively reinforcing arms race with the metrics. Goodhart's law should apply to baskets of metrics just as much as to any single one. This should be so because the only way to *deploy* any metric as an evaluation standard without violating Goodhart's law is to keep the metric or standard secret. The basket approach only does this by being too complex to understand, or proprietary and thus secret, so that it is *de facto* hidden from the day-to-day practice of individuals subject to performance evaluation. That strategy will have only temporary success because academics are clever and are paid to solve puzzles of this kind, hence the inevitability of an arms race. Moreover, research production systems and research evaluation systems are inextricably linked by processes such as peer review for publication and grant award, which must be transparent (or would be extremely unethical, like not telling assistant professors what is required for tenure). So I don't see an end game in the basket of metrics approach, other than mutually assured destruction. Now that may not be a bad thing. Creative destruction of an old-fashioned, biased, elite research system might be justified if it brings down an ill-suited, ill-fitting, conservative system of research evaluation with it.

However, there is more than one way to creatively reimagine a role for judgment alongside metrics in performance evaluation. I return to Goodhart's law as a way of talking about how to reframe the question. A reframed question may lead to different kinds of solutions than a basket approach or *Leiden's* policy demand that inappropriate measures such as impact factors be dropped as standards for individual authors.

Goodhart's law teaches that to understand what goes on in the world of metrics-based academic performance evaluation, we should look beyond how metrics might be gamed and beyond the question whether, because the metrics are used as standards, gaming the metrics is a form of misconduct in the bad sense of gaining something undeserved through the evaluation process. What we need to understand is how what individuals *do* in their research production processes in response to the standard changes in relation to the metrics as a consequence of their use as standards. That is not merely a matter of assessing whether, how, and to what extent researchers from this or that part of the globe decide to submit their research findings, data, figures, or code to a "predatory" or "junk" journal rather than to a "legitimate" one—nor whether they decide to analyze other people's hard-won published data rather than go to the trouble and expense of generating their own data, nor to Photoshop old figures rather than produce new ones based on new data. It is a matter of understanding deeply, and on a fine-grained scale, how research production—the whole content of research activity, and not just "publishing"—is changing in an environment where performance metrics function as standards.

Assessing the content of research *activity* is not the kind of problem that can be solved with more data captured in metrics with the digital discovery methods discussed in this volume. Transformations of research production processes at the level of the conduct and decision making of individual scientists and small teams in their day-to-day workflows can only be captured by good old-fashioned social science *research* that social scientists use to figure out what, how, and why people do what they do: interviews, participant observation, close readings of unpublished and published work, surveys, and now distance communication through the internet.

The strategic intervention I propose is this: Let's assume Goodhart's law is true. If it's true, we should *expect* "gaming" and "mis-"conduct in the system as a normal or typical part of the workflow of any well-adapted research production system whenever metrics are used as standards of research performance evaluation. So, it may be more fruitful to look for situations in which research performance appears *not* to adapt as a means of revealing *breakdown* of the kinds of behaviors one should expect to find when metrics are made into standards.

Adaptation is the most plausible state of affairs in response to the imposition of the social forces represented by metrics used as standards. In that light, the phenomenon of "predatory" or "junk" journals is one kind of response we might consider interpreting as "mis-" rather than "bad" conduct. Maladaptation or nonadaptation, in the sense of nonconformity or noncompliance to a standard, *should* appear anomalous or "bad" to an institutional policy regime that expects compliance. More radically, hacking metrics as a form of experimental, *transparent* mal- or nonadaptation might serve as a means of understanding the mechanisms and dynamics of adaptive responsiveness to imposition of metrics as standards. The further point is that we can think about two ways of understanding what Goodhart's law tells us about the variety of kinds of behavior subject to lumping under "gaming the system." One is the sort of gaming discussed in this volume, and it is quite interesting.

The other kind of possible response to Goodhart's law is to *embrace* gaming the system as a tool for experimental intervention into research *production* systems, extending the traditional observational tools of social science research. The idea is to make hacking the research system part of a research *program* for understanding what causal consequences, at the micro-level of research production, follow from the imposition of social forces at the macro-level of social organization. (In a sense, "predatory" or "junk" journals can be viewed as leading the way in hacking, provided their interventions are interpreted according to the proper causal-experimental framework. We can learn much from their tactics even if we eschew their profit motives.) To use hacking as a research tool, we need to adapt research production work *on science metrics* to a new goal: experimental intervention into research performance systems that are subjected to metrics-based performance evaluation. To do that, we would need to create an artificial (i.e., experimental) publication system in which such research work could be published and an artificial (i.e., experimental) research specialty that organizes it.

We need not only to multiply the *metrics* as in the basket approach, but also to actually hack gaming *behaviors* in order to find out how the causal structure of individual researcher behavior is changed by the imposition of metrics. Traditionalists might wonder why we couldn't just interview people who may have been subjected to behavior change in the face of the metric tide or do longitudinal studies of people experiencing different metrics environments. We could do this of course. But it seems unlikely we would observe appropriate contrasts among scientists in their experiences of metrics-as-standards to develop much insight into causes. Even comparison of researchers in closely related specialties or in different national contexts or across historical periods would have so many confounding variables at work as to render them of limited utility for discovering causal impact.

What I propose instead would be a program of intervention with individual researchers and research groups, while historians, sociologists, anthropologists, and philosophers—such as science studies researchers—study them, in an enhanced environment where science studies takes science metrics into *account* descriptively while also manipulating the research environment in which performance is evaluated. Such experiments are taking place within the current research system, for example, by Labbé's experiments with his fictional author, Ike Antkare (this volume, chapter 14). I propose parallel experimentation with the research publication system itself, using methods inspired by Labbé and others.

The tension between measurement and standards deriving from Goodhart's law, as I've noted, is that because humans are reflexive, metrics used as standards have to be kept secret, otherwise the people who are measured will change their behavior in ways that defeat the value of the metric as a measure. On the other hand, a standard has to be transparent: it is unfair and unproductive to hold people to a standard they cannot strive to meet. The problem is that secrecy and transparency don't work so well together.

I propose we use our own reflexivity as technoscience researchers and scholars of the operation of academic metrics, in the tradition of medical self-experimentation, to manipulate experimentally the conduct of science studies research to find out what kind of changes can be brought about by exploiting and manipulating metrics-based measurement of performance.

In other words, a way Goodhart's law could turn out to be *true* is not merely that people are responsive to these forces in ways that change the causal structure of their behavior—Goodhart's law could be true because people respond reflexively to satisfy Goodhart's law *on purpose* as a way of playing a game. Just as video game hackers improve game play by intentionally violating the designs of the game designers to make the game play differently, thereby inventing a new game, science studies research might engage in experimental manipulation of their own metrics as a means of understanding contemporary science in the age of metric tides.

That would be to change the *causal* relationships of the game—to make it a different game, not merely to play "the game" by explicitly engaging in "mis"-conduct or "bad" conduct with respect to the institutionalized rules of the game, but to invent a new game by hacking the old one. What can we do to investigate metrics on a par with the kind of hacking that goes on in the video game world?

One way to do it is to create a collection of journals designed to publish research on science metrics but which includes in their mission the explicit "gaming" of metrics that are used as standards in performance evaluation. Call them "PuLP" for "Public Library of Philosophy" on the model of PLoS—Public Library of Science. (Thanks to Jonathan Eisen for suggesting the acronym PuLP.)

The mission statement for "PuLP-ONE" would include:

1. Reviews and surveys of current performance measures/metrics and which ones are used by whoever's policies as standards. As it happens, journals are already beginning to appear that have this scope, for example, *Research Integrity and Peer Review* (http://researchintegrityjournal.biomedcentral.com).
2. Success and failure impact narratives authored by individuals and groups about what metrics have done for/to them.
3. Science studies research on practices considering the role of science metrics in the conduct of research or its evaluation, such as the chapters in this volume.
4. Overt hacking of metrics by publishing work of the above three kinds and of any other kind (including machine-generated papers) in whatever nominal field of study so as to explicitly and transparently attempt to manipulate metrics and thus to game standards.

A system of PuLP journals, a public library of hacker science studies, so long as it is sufficiently amusing to indicate transparently that its goals are not business as usual, would be designed primarily to transparently and openly *intervene into* how metrics affect research or behavior. PuLP would, for example, publish science studies work with, say, five hundred authors, citing articles in other PuLP journals and also other journals the authors regularly publish in. By manipulating the number of articles published in a particular PuLP journal and the number of citations to articles in that journal, we could not only manipulate the Journal Impact Factor, as many "predatory" journals already do, we could also engineer whatever impact factor we wanted, showing just how arbitrary a measure it is and just how irrelevant to the content of research. The "h-index," as a measure of author impact, could be manipulated by engineering many citations to works published by that author from other PuLP journals as well as by seeking agreement of those publishing in PuLP to cite PuLP journals in their works published in non-PuLP ("civilian") journals. The full range of tactics discussed in other chapters in this volume would not only be available to authors in PuLP, but would also be part of the mission to use these methods and build new ones.

Because the journals would also be designed to publish scholarly research *assessments* and *interpretations* of how changes in research production behaviors undermine metrics and at the same time reorient or redirect workflows and production of whole fields of scientific research, PuLP journals could not be dismissed as "mere" junk, especially if their papers published under missions one through three are of high quality. PuLP journals would provide a "respectable" outlet for science metrics research. If this mission is held to community standards of scholarship, it would be harder to discount the hacker work out of hand on the self-serving grounds usually supporting metrics-based assessment in the first place: appearance in journals that meet metrics-based standards.

Scholarly publications in PuLP journals that report, assess, and interpret responses to metrics-based research performance and impact narratives could serve as a basis for designing, announcing, and conducting new hacking techniques and experiments, as the project would presumably kick the arms race into a higher gear, especially when hundreds of science studies researchers begin to submit experimental CVs for personnel evaluation with dozens of publications per month.

The primary mission of PuLP would be to undermine existing metrics by embracing and exploiting Goodhart's law. PuLP-ONE would be a *transparent* journal for hacking the metrics—not designed for the sake of gain like a junk journal might be: to make money or earn prestige for authors, but for the sake of understanding experimentally how metrics manipulate social behavior, thereby showing how they are subject to gaming and to undermine their use as standards lacking a balanced involvement of judgment in evaluation. The goal would not be just to tell stories about how scientists conduct their research and the sorts of pressures they experience, but also to interrogate and ultimately change practices of research performance evaluation.

In providing a forum for *review* of research metrics and *assessment* of responses to metrics as standards, and thus the material platform needed to design hacker interventions, PuLP might help end the arms race of metrics and gaming by revealing the likely decline and fall of the research system from the inadvertent, unintended consequences of continued pursuit of metrics-based evaluation.

Enthusiasm for metrics is reinforced by bigger and bigger data, so it is probably necessary to do this experiment and not only talk about it. The aim is not to do away with research metrics but to return the project of *evaluating the metrics* to serve researcher valuation of the content of their research and to repair the damage caused by diverting this value into a side effect of a transformed research system that mainly values the advancement of auditable knowledge. If sustaining the research enterprise requires hacking the metrics, let the games begin!

References

Chamayou, Grégoire. 2009. "Petits Conseils aux Enseignants-Chercheurs qui Voudront Réussir Leur Évaluation." *Contretemps*. February 24, 2009, http://www.contretemps.eu/petits-conseils-enseignants-chercheurs-qui-voudront-reussir-leur-evaluation/. Consulted June 16, 2017. Read from Google Translate into English.

Daston, Lorraine, and Peter Galison. 1992. "The Image of Objectivity." *Representations* 40:81–128.

Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. Brooklyn, NY: Zone Books.

Dumit, Joseph. 2012. *Drugs for Life: How Pharmaceutical Companies Define Our Health*. Durham, NC: Duke University Press.

Galison, Peter. 1998. "Judgment Against Objectivity," in C.A. Jones and P. Galison (eds.). *Picturing Science, Producing Art*, 327–359, New York: Routledge.

Goodhart, C. A. E. 1975. "Monetary Relationships: A View from Threadneedle Street," in *Papers in Monetary Economics*, Volume 1, Reserve Bank of Australia. Quoted in: *Goodhart's Law: Its Origins, Meaning and Implications for Monetary Policy*, by K. Alec Chrystal and Paul D. Mizen. Prepared for the Festschrift in honor of Charles Goodhart to be held on November 15–16, 2001 at the Bank of England. November 12, 2001, http://cyberlibris.typepad.com/blog/files/Goodharts_Law.pdf.

Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. "The Leiden Manifesto for Research Metrics." *Nature* 520(April 23):429–431.

Lucas, Robert 1976. "Econometric Policy Evaluation: A Critique," in K. Brunner and A. Meltzer (eds.). *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy 1, 19–46, New York: American Elsevier.

Power, Michael 1997. *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Star, Susan L., and James R. Griesemer. 1989. "Institutional Ecology, 'Translations,' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939." *Social Studies of Science* 19:387–420.

Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Steven Hill, Richard Jones, Roger Kain, Simon Kerridge, Mike Thelwall, Jane Tinkler, Ian Viney, Paul Wouters, Jude Hill, and Ben Johnson. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. doi:10.13140/RG.2.1.4929.1363.

II

Collaborative Manipulations

Collaborative Manipulations analyzes how the pressure to "have impact or perish" helps to spread gaming practices on a global scale while also turning them into collaborative and even institutional practices. The dissemination, institutionalization, and increasingly collaborative nature of these gamings create unforeseen consequences, including the remarkably innovative and ever-changing nature of these manipulations. In addition to the peer review and citation rings mentioned above, these trends may be exemplified by researchers who collaborate with pharmaceutical industries to lend their names to ghostwritten articles in order to "harvest" their numerous citations without lifting much of a finger, or, in other cases, by universities that provide cash bonuses to professors who publish in top journals, which in turn help the university climb in the international rankings.

Barbara Kehm discusses why universities dedicate significant resources to improve their global university ranking, and how a high-ranking university may come to be treated as a nation's "indicator for the scientific and technological capacity and productive efficiency." As governments often make educational reforms and funding decisions based on the global rankings of their universities, we see how the compounded effect of impact-seeking publication strategies by many individual authors trickles up to the institutional and then national level, eventually causing global effects—and gaming opportunities.

In a similar vein, Sarah de Rijcke and Tereza Stöckelová argue that the European research policies' focus on a publication's "international impact" as a stand-in for quality ends up reinforcing hierarchies between the "international" North and the "parochial" South. This problematic divide, however, can lend itself to gaming. Because universities' rankings benefit from prolific faculty who publish internationally, these institutions tend to be slow disciplining the faculty who have published in journals that, although possibly originating from the Global South and appearing in the now-defunct Beall's list of "predatory journals," can still be counted as "international." Universities often turn an equally blind eye, for very much the same reason, when their faculty list fictional "international" co-authors on their publications, or boast to serve on editorial boards of questionable but "international" journals.

Focusing on a different type of publication/academic pressure, Daniele Fanelli makes the convincing argument that co-authorship is, effectively, a kind of collaborative gaming, or at least a gaming that hinges on collaboration. In response to publication pressures from the university, scientists co-author articles with more and more authors so as to match higher productivity benchmarks. Rather than "salami-slicing" their research into multiple publications, scientists "salami-slice" their collaborations. As papers become increasingly co-authored, individual scientists can list more papers on their vitae.

Sergio Sismondo examines a different kind of "authorship gaming," one with two players driven by two very different but compatible sets of goals. Pharmaceutical companies attribute ghostwritten articles to willing and influential scientists—a practice that gives the article the "veneer of having been written by independent researchers, instead of by a coordinated industry team." This allows pharmaceutical companies to publish articles that are carefully crafted to advertise their drugs and yet appear under the guise of legitimate scientific publications. The "authors," on the other hand, receive credit for publications they have not authored but that, due to their high production quality and the professional handling of submissions and revisions by professional writing companies, have a higher acceptance rate, shorter time to publication, and more citations than comparable articles written by independent academics.

6

Global University Rankings: Impacts and Applications

Barbara M. Kehm

The Ranking Game

In December 2011, the journal *Science* published the information that two Saudi-Arabian universities were massively recruiting highly cited research stars from Cambridge, Harvard, and other universities who had made it onto the Institute for Scientific Information (ISI) list of most frequently cited researchers. For about $70,000 per year, they were offered an affiliation to these universities in exchange for the obligation to be present once a year for a short time and to indicate in all their publications their affiliation to the respective Saudi-Arabian university. The result was that within two to three years both institutions made it from not listed at all into the group of the top two hundred to three hundred in the Shanghai Jiao Tong Academic Ranking of World Universities (ARWU). Thus universities are buying the reputation of researchers in order to increase their own reputation. Not all researchers who were contacted could be bought. However, in March 2012, the largest Australian newspaper, *The Australian*, published a list of sixty frequently cited researchers who had been appointed as "distinguished scientists" at one of the two Saudi-Arabian universities. Altogether the list comprises a number of researchers from top universities in the United States, Canada, Europe, Asia, and Australia. All of them are men, and some are already retired.

In 2012, the Australian University of New South Wales published a job advertisement for "Strategic Reputation Management" and the Australian La Trobe University was looking for a "Manager for Institutional Rankings." For an annual salary of $100,000, the job descriptions comprised among other things the task to manage the university's relationships to ranking agencies and to "maximize" or "optimize" the respective institution's ranking position (Inside Higher Ed, March 22, 2013). In the same article, University of New South Wales' pro-vice chancellor was quoted to have stated that it was essential for a university to have a team that takes care of the proper presentation of the numbers.

But does this kind of manipulation work? And more importantly is such a practice still related to good science and scholarship? It becomes clear that rankings seduce and coerce at the same time. Those universities that want to participate in the ranking game have to internalize and institutionalize the logic of rankings. Morphew and Swanson (2011) have pointed out that "rankings determine and even codify which kinds of organizational behavior and practices are legitimate." Therefore, the players know that they have to be successful under the conditions of the measurements. Ranking positions have a signaling effect and contribute in a seemingly objective way to the discussions about what constitutes quality in higher education. Thus, universities use a number of gaming techniques in order to improve their ranking positions. Morphew and Swanson (2011) provide further examples from US universities: adjunct instructors are not counted when reporting the percentage of full-time faculty employed; admission data are presented in such a way that they signal a high level of selectivity; law schools are spending high amounts of money for glossy brochures to influence reputation scores.

Accordingly, the authors come to the conclusion that these forms of participation in the ranking game simultaneously challenge and reinforce the legitimacy of rankings. A classical paradox!

In her survey among university leaders published in 2007, Ellen Hazelkorn found that ninety-three percent of the respondents wanted to improve the position of their university in national rankings and eighty-two percent wanted to improve the position of their university in international rankings. Seventy percent wanted to see their university among the top ten percent in national rankings, and seventy-one percent wanted to see their university among the top twenty-five percent in international rankings. However, other studies have shown that variations in ranking positions are only temporary and mostly disappear after two years. Between 1988 and 1998, twenty universities out of the top twenty-five identified by the *U.S. News and World Report* ranking never fell out of this top group. Therefore, it is almost impossible for other universities to move into this group. It seems, however, that the multitude of specialized rankings that have been developed in recent years (e.g., top universities under fifty years of age, etc.) is, at least in part, also meant to help these other universities to make it to some kind of "top."

Global rankings like the ARWU ranking of Shanghai Jiao Tong University or the ranking produced by the Times Higher Education provide information about four to six percent of all universities globally. As a consequence, "all universities are judged on the basis of criteria that are only appropriate for top universities" (Rauhvargers, 2011). This leads to the construction of a "deficit model" (Locke, 2011) that drives all universities that participate in the ranking game into a perpetual race to improve their ranking position. At the same time, rankings offer hardly a possibility to rise into the top group. So, why all the excitement then?

First, good ranking positions trigger the famous Matthew effect. Better students and academics apply, donations by alumni rise, and, in many countries, such universities receive increased budget allocations by the state. Second, rankings distribute reputation. And reputation is an important immaterial resource, difficult to build up and easy to lose. Third, rankings are popular among political decision makers—on the one hand, because they reduce complexity, and, on the other hand, because high ranking positions of one or more universities in the country have become an indicator for the scientific and technological capacity and productive efficiency of the national economy as such.

But rankings do not provide any information about the quality of a university as a whole, even if they pretend to do just that. And there are only few players that have the capacity to play the game profitably. According to Salmi (2009), these are, in particular, large, preferably older and research-intensive universities with a broad spectrum of subjects (i.e., including medicine) located in the English-speaking world. In addition, they have to have three further features: abundant resources, a benevolent management, and a concentration of talent. Other potential players should better abstain from playing the game because it might lead to problematic management decisions.

Examples of Resistance

In the meantime, rankings have multiplied at national as well as international levels. Hazelkorn (2011) identified altogether nine active global rankings and more than fifty national rankings. And the number has probably risen by now. Despite the fact that many experts have argued that rankings are here to stay and the task is to improve them rather than ignore them, resistance against rankings has started and it is coming from the academic side. Without being able to provide a complete overview, just a few examples should suffice: the Australian James Cook University is ignoring the ARWU ranking; some universities in the United States, among them the prestigious Annapolis Group, are boycotting the *US News and World Report* ranking either as a whole or its reputation survey part; and a number of Canadian universities have refused to participate in the Maclean's University Ranking.

In Germany, several learned societies have by now recommended to boycott the ranking carried out by the Center for Higher Education Development (CHE), among them the German Society for Sociology, the German Society of Historians, the German Society of Chemists, and the German Society of Education. These organizations have issued appeals to both their individual academic members and the respective university departments not to submit any data to ranking agencies. In addition, four universities have announced not to submit any data for purposes of rankings: Hamburg, Leipzig, Cologne, and the Distance University of Hagen. The view of these institutional ranking opponents is that the generation and proper presentation of data for the CHE ranking would require the work capacity of more than ten full-time employed people, and they were not prepared to finance this any longer when the task of a university is to provide a good education to the students.

In March 2013, three hundred economics professors in Germany rebelled against a ranking of business studies and economics professors carried out by the *Handelsblatt*, a daily newspaper focusing on economic news. Their main argument was that such a ranking worked with wrong incentives, and that headings like "Germany in search of the super prof" were getting too tacky. For readers who are not very familiar with Germany, there is a German television show called "Germany in Search of the Super Star" in which young talents (mostly singers) compete against each other. The show became known in particular for its prejudiced and mean comments by the jurors.

In the last part of this contribution, a few thoughts are offered about why rankings have met the resistance of academics but are loved by policy makers (and frequently university leaders as well). It is also an attempt to provide a more theoretical framing for the phenomena that have been described so far.

Rankings as a Form of Transnational Policy Coordination

It is an interesting phenomenon that rankings have become rather important for national policy makers and institutional leaders but have met with resistance from the academic side. This is not the place to go into the criticism of methodological flaws, the bias toward English language publications, the focus on research only, and other well-known critical aspects. It is more interesting to discuss the ways in which the phenomenon of rankings has been theoretically framed.

Erkkilä (2013) has framed rankings as a policy instrument of global university governance, and others have analyzed it as a form of transnational policy coordination. What has been observed is that the outcomes of rankings constitute a policy problem at the national as well as, for example, the European level, which has led to policy changes. Although the ARWU ranking originally was a domestic policy instrument in order to evaluate how Chinese universities fare against top universities in the rest of the world, its outcomes have created a global narrative of higher education competition, which itself is used as an indicator for the competitiveness and strength of national and (in Europe) regional economies. Thus we have a double transfer to meta levels. Rankings have become a symbol of economic status because it is argued that the more universities in a given country or region are ranked among the top ten, fifty, one hundred, or five hundred, the higher is the economic reputation and innovative capacity of that country or region. And, as Erkkilä argues, despite the fact that global rankings do not possess a norm-giving authority, they have influenced policy decisions. In Germany, they triggered the “excellence initiative,” and at the European level, they contributed to the decision of funding the U-Multirank Project. And this has led to another paradox, namely that global rankings address individual higher education institutions while at the same time having geographical implications (i.e., German versus British universities or European higher education versus US higher education). This contributes clearly to isomorphism in national policy making and institutional leadership despite the calls for institutional diversity.

The ARWU ranking became the start of a global assessment of higher education that linked to new forms of global and transnational governance building on comparison and evidence-based decision making. Basically the outcomes of the ranking served as the evidence policy makers needed in order to introduce reforms and overcome resistance. What we have here is actually the governance of complexity in the face of globalization. Thus, global rankings can be understood as a “transnational policy script” (Gornitzka, 2013) that has diffused into different national contexts and has become a reference point for legitimizing higher education reforms. Using examples from Germany and the European level again, the “policy script” was translated in Germany into giving up the traditional legal homogeneity with which universities were treated by the state and introducing competition, while the "policy script" was translated at the European level by establishing a "modernization agenda" for European higher education.

Holzinger and Knill (2005) have described the process of transnational policy coordination as a form of transnational communication leading to policy diffusion. This transnational communication is characterized by four mechanisms: 1) lesson drawing; 2) transnational problem solving; 3) policy emulation; and 4) international policy promotion.

Lesson drawing is a process where states learn from each other what can be done when problems occur. It implies the existence of "best practice," which is taken as an efficient way to reform policies by using examples and models developed elsewhere. In transnational problem solving, solutions are sought and found in transnational networks or epistemic communities that—with the help of transfer agents like international organizations—facilitate the exchange between polities and spread the policy. Policy emulation is a one-directional policy transfer that basically consists of copying and implementing a policy without adaptation to local, regional, or national contexts. Thus, policy emulation is imitation rather than innovation. In international policy promotion, finally, we have specialized organizations that actively promote certain policies while defining objectives and standards in an international setting.

It can be argued that the spread of rankings as an instrument of transnational policy coordination consists of a mixture of transnational problem solving and international policy promotion. Increasingly there are groups of academics involved in rankings, the best example being the European U-Multirank Consortium (see https://www.umultirank.org), which is funded by the European Commission. It advocates and supports the idea of developing a European university ranking and thus acts as an agent for the promotion of such a policy in Europe.

Conclusions

If we look at the history of rankings, we can observe that they started out as an academic exercise focusing on disciplines or units rather than whole institutions. The views vary about the beginning of rankings. Dill (2009) identifies the first ranking as the one that was carried out in 1925 by Raymond Hughes, a professor of chemistry and later vice-chancellor of Miami University. Hughes did a reputation survey of graduate programs.

Hazelkorn (2011) dates the first ranking earlier, namely to the year 1910, by referring to James Catelli, a US psychologist and professor at the University of Pennsylvania. And Salmi and Saroyan (2007) observe first ranking attempts from 1870 onward when a commission of the US Bureau of Education began to publish annual statistical reports that also included a classification of institutions.

But while the first rankings in the United States were mostly carried out by active academics, the first *U.S. News and World Report* ranking from 1983 was a commercial ranking that ranked whole institutions. This triggered an imitation frenzy by other weeklies and dailies in order to increase their sold copies. To name just a few, we have the *British Times Higher Education* and the British *Guardian*, the German *ZEIT* and *Spiegel*, the French *Nouvel Observateur*, the Irish *Sunday Times*, the Italian *La Repubblica*, the Russian *Finance*, the Canadian *Maclean's*, and probably many others.

The ARWU ranking demonstrated the beginning of a reappropriation of rankings by academics, and we have currently more rankings that are carried out again by academics. But the impacts and political uses of rankings have changed. Rankings are used as a policy instrument for what is nowadays called evidence-based political decision making. Ranking results present a simple, although undercomplex (i.e., not appropriately reflecting the actual complexity of what universities are about), hierarchy expressed in a positional number according to which funding can be allocated and legitimized by governments. Thus rankings establish a deficit model (Locke, 2011) according to which no institution is ever good enough, except the one on the top, or let's say the few on the top. This triggers a race for position that disregards issues of quality improvement and diversity of mission. In other words, rankings seduce and coerce at the same time (Locke, 2011). By now, every national government wants at least "one Harvard University" in their country in order to demonstrate to the world that it is economically competitive. And thus the ranking results become themselves indicators or, more exactly, proxies for something else, and national governments might make decisions on the basis of the symbolic value of ranking scales. This is a truly postmodern phenomenon. The positional hierarchy of universities created by rankings makes the actual reality of universities and what they are about disappear. The hierarchy is then shifted into the economic sphere of nations or regions, thereby constituting a decontextualized symbolic value that itself can be charged with new meaning and thus creating a new material reality that is no longer related to its original.

References

Dill, D. 2009. “Convergence and Diversity: The Role and Influence of University Rankings.” In: Kehm, B. M., and B. Stensaker (Eds.). *University Rankings, Diversity and the New Landscape of Higher Education*. Rotterdam, Taipei: Sense Publishers, pp. 97–116.

Erkkilä, T. 2013. “University Rankings and European Higher Education.” In: Erkkilä, T. (Ed.). *Global University Rankings. Challenges for European Higher Education*. Basingstoke: Palgrave Macmillan, pp. 3–19.

Gornitzka, A. 2013. “Channel, Filter or Buffer? National Policy Responses to Global Rankings.” In: Erkkilä, T. (Ed.). *Global University Rankings. Challenges for European Higher Education*. Basingstoke: Palgrave Macmillan, pp.75–91.

Hazelkorn, E. 2007. “The Impact of League Tables and Ranking Systems on Higher Education Decision Making.” *Higher Education Management and Policy*, Vol. 19, No. 2, pp. 1–24.

Hazelkorn, E. 2011. *Rankings and the Reshaping of Higher Education. The Battle for World-Class Excellence*. Houndmills, Basingstoke, and New York: Palgrave Macmillan.

Holzinger, K., and C. Knill. 2005. “Causes and Conditions of Cross-National Policy Convergence.” *Journal of European Policy*, Vol. 12, No. 5, pp. 775–796.

Inside Higher Ed. 2013. March 22. https://www.insiderhighered.com.

Kehm, B. M., and T. Erkkilä (Eds.). 2014. “Global University Rankings. A Critical Assessment.” *European Journal of Education*, Vol. 49, No. 1, March.

Locke, W. 2011. “The Institutionalization of Rankings: Managing Status Anxiety in an Increasingly Marketized Environment.” In: Shin, J. C., R. K., Toutkoushian, and U. Teichler (Eds.). *University Rankings. Theoretical Basis, Methodology and Impacts on Global Higher Education*. Dordrecht: Springer, pp. 201–228.

Morphew, C. C., and C. Swanson. 2011. “On the Efficacy of Raising Your University’s Ranking.” In: Shin, J. C., R. K. Toutkoushian, and U. Teichler (Eds.). *University Rankings. Theoretical Basis, Methodology and Impacts on Global Higher Education*. Dordrecht: Springer, pp. 185–199

Rauhvargers, A. 2011. *Global University Rankings and Their Impact. EUA Report on Rankings*. Brussels: European University Association.

Salmi, J. 2009. *The Challenge of Constructing World-Class Universities*. Washington, DC: The World Bank.

Salmi, J., and A. Saroyan. 2007. “League Tables as Policy Instruments: Uses and Misuses.” *Higher Education Management and Policy*, Vol. 9, No. 2, pp. 31–68.

Yudkevich, M., P. G. Altbach, and L. E. Rumbley (Eds.). 2016. *The Global Academic Rankings Game: Changing Institutional Policy, Practice, and Academic Life*. New York: Routledge.

7

Predatory Publishing and the Imperative of International Productivity: Feeding Off and Feeding Up the Dominant

Sarah de Rijcke and Tereza Stöckelová

"Be international!" This imperative can hardly be overlooked in current European research policy and research evaluation.[1] The imports of "internationalization" manifest prominently in how particular value is attached to "international visibility," "international impact," or the international character of publication venues. The international is used as a trope on a number of levels: in EU funding schemes, in project goals that guide national assessment exercises, in output measurements, in the formulation of institutional research missions, and in tenure-track criteria. Especially in smaller countries, such as the Netherlands and the Czech Republic, the international is often taken as an unquestioned proxy for quality, proving recognition of value and impact beyond the "academic pods." Consequently, the international, the national, and the local constitute a clear normative hierarchy. For example, it is taken for granted that international excellence encompasses national excellence and (as such) is supposedly more valuable.

Inspired by Lin and Law's discussion of "modes of international" (2013, 2014), we argue in this chapter that gaming metrics, predatory publishing, and exploiting the model of gold open access (Beall, 2012) can be partly understood as a logical response to the imperative of internationalization going wild. It enacts a different, yet dubious, alternative mode of internationalization for those researchers and institutions who fail—for better or worse—within the established mode of international, with its epistemic and economic centers in the global, Anglophone North/West. In this chapter, we zoom in on a recent misconduct case in the Czech Republic to show how the imperative of internationalization and productivity inscribed in the country's research assessment framework impinges on institutional and individual publication strategies and produces a market for gaming in the academy.

Taking the Imperatives to the Extreme

In 2015, a major debate on publishing and research evaluation was opened up in the Czech academy. It was provoked by controversy over a highly productive junior researcher at the Faculty of Social Sciences of Charles University. At first sight, he might look like a paradigmatic case of a successful scholar with a long list of international publications, collaborations, and co-authorships—exactly what the current research policy in the Czech Republic holds as a normative ideal. However, on second sight and when some of his colleagues from the department started to closely scrutinize his production, the case turned out to be something significantly different: a sophisticated attempt to game the current research assessment system on various levels—or rather, to take the imperative of the system to the extreme by some perfectly legitimate and some less legitimate ways. To understand what happened, let us first briefly describe the genesis and current state of research assessment in the country.

The post-1989 changes in the Czech Republic concerned not only political and economic institutions, but also academic ones. One of the most fundamental changes was the establishment of the Czech Grant Agency in 1993 and the introduction, in various forms, of competitive funding of academic research. In 2001, and largely from the initiative of a few natural scientists who came back to the Czech Republic in 1990 after spending several years in the West, the first version of a new methodology for the quantitative assessment of institutional-level research performance was introduced. Its impact on research funding of academic institutions and the "value" imputed to individual scholars has since then gradually increased. The central building blocks of the evaluation methodology are so-called RIV-points (RIV standing for "Information Register of R&D results"[2]), assigned to predefined types of outputs (including journal articles, monographs, patents, and prototypes) and meant to reflect their academic and user value (Office of the Government of the Czech Republic, 2013).[3] One of the key claimed rationales of the evaluation methodology was to create an objective "machine" that would increase the transparency of the research system and depoliticize its governance. However, during the last fifteen years, the methodology developed into a convoluted metrics-based amalgamation with many unclear algorithms and weights that are far from transparent, not only for "ordinary" researchers, but even for research policy managers at the national level (Miholová and Majer, 2016). At present, the evaluation methodology's criteria for "quality recognition" soak through the entire system. They have a significant—even if, at times, indirect—impact on academic hiring and promotion procedures, individual research grant endowment, and the funding allocation of public research institutions.

A key trope of the research policy reforms since the 1990s has been internationalization, and this trope is also inscribed into the current evaluation methodology. This is understandable in a small country where many disciplines tended to operate in closed circles consisting of local scholars. However, it is more problematic that the international oftentimes stands as a value in itself—unquestioned and undisputed, for example, there is currently nearly no peer-review evaluation of journal articles within the national evaluation framework (a peer-review evaluation of a limited number of outputs submitted by research organizations as "excellent" was introduced in 2015) and the journal impact metrics provided by Web of Science (WoS) and SCOPUS are taken for granted as proxies for international recognition and quality. This is the context in which junior academics start to build their publication record and careers.

We now return to the controversy. Having gained his PhD in 2007, the academic in question has claimed to have co-authored or co-edited seventeen "scientific monographs" between 2011 and 2013 and more than eighty journal articles between 2006 and 2015.[4] Apart from the extreme productivity, four aspects of his CV are noteworthy. Firstly, the author also acts as an editor in chief, editorial board member, and even publisher of some of the "European" or "international" journals listed on his CV.[5] All these journals are English language, target an international audience, and have an international review board and international pool of authors. Secondly, even if in SCOPUS, some of the journals on his publication list were also listed in Jeffrey Beall's database of predatory journals.[6] Thirdly, some of the co-authors on these articles in predatory journals were colleagues from the faculty—including the current head of the department.[7] And, finally, as the author later confirmed, one of his co-authors was discovered to be a fictional character supposedly affiliated with prestigious Western European universities (first the University of Strasbourg and later the University of Cambridge).[8]

While some of the academic's actions were rather extreme, or even "crafty" (e.g., the invented co-author; see also Marie-Andrée Jacob's chapter on template, dexterity, and publication ethics), we have to acknowledge they have definitely been in line with the current imperative of internationalization. The researcher tried to gain "Western" recognition and certification (listing on the WoS and SCOPUS databases) for his publishing activities as an author, editor, editorial board member, and publisher based in the East. Interestingly, he not only strove to gain a position in the existing international playing field (which is what the research policy framework in fact tries to encourage), but also, as a skillful academic entrepreneur, to rework and reorder the field at one go by creating new journals and forging new East–West alliances (even if at times with fictitious co-authors). He also specifically offered his teaching and publication "services" to researchers from Russia and Eastern Europe in relation to whom he positioned his activities as international. Apparently, he aimed at the enactment of a different international than the one of current global science, in which the international in fact equals the West. While in general we might have some sympathy for attempts at destabilizing the global asymmetry (Stöckelová, 2012), his means and ways of doing so are rather problematic.

As a result of a major controversy at the faculty level, during which "whistle-blowing" colleagues from the department filed a complaint to the Ethical Commission of Charles University (the complaint was deferred[9]), and following the publication of a number of articles in national public media, the author's contract was terminated in September 2015. In response to the increasing media and academic community pressure, the faculty openly distanced itself from unethical publishing practices connected with the case. It issued "publication rules" that warned against predatory journals and vanity press publishers, such as the well-known vanity press Lambert Academic Publishing, in which over twenty "international" monographs of the faculty members had been published since 2010.[10] Some other faculties and universities in the country followed suit.

Interestingly, the "international" standards for quality assessment did not seem to count equally for all involved. Though playing the game led to several promotions for the researcher who was later accused of misconduct, when push came to shove, the same rules did not apply to the key whistle-blower, though he and his research group were doing quite well by these standards.[11] Debatably, a few weeks later, after the termination of the perpetrator's contract, the contract of the main whistle-blower was not renewed either—in spite of wide support for his actions from the social science community.[12] The faculty chiefly adhered to a "bad apple" approach, a relatively common strategy in misconduct cases in the sense that measures are often taken mainly at the level of individuals.

Cui Bono?

Calls for more transparent, trustworthy quality control mechanisms and more open infrastructures for communicating and publishing research are currently widely heard in European science policy. The European Commission has introduced several framework programs that focus in particular on responsible research and innovation, and on "open science." In 2020, all scientific and scholarly output should be freely available by way of open access. Another important aim for 2020 is a fundamentally novel approach to data (re)use, based on open data models. But change will not come easily, with vested interests of established academic elites and large commercial actors with their entrenched infrastructures for publishing and evaluating research. Paradoxically, part of the answer seems to lie in the hands of exactly these commercial parties. At present, they appear to be the ultimate gatekeepers of the "international." The critique of predatory journals inadvertently makes a very strong case for the value added by corporate, indexed outlets and black-boxed, commercially endorsed algorithms. Predatory journals seem to play right in the hands of corporate publishers as a confirmation of the dangers of uncontrolled open access.

At the same time, the predatory publishing industry managed to develop a business model that taps into both the "open science" *and* the "commercial" publishing models and normative frameworks. Evidently, some of the appeal of predatory journals and vanity publishers lies in their offering cheap, accessible vehicles for the "international"—certainly when compared to the costlier "gold" open-access publications, with quality control and more or less US- and Eurocentric gatekeepers. Also, the predatory publishing business model closely mimics and reproduces the standards and incentive structures of the "global," dominant publishing industry. This is an industry in which the journal and the journal article are the most valuable means of communication for international recognition and visibility, within a "market world of justification" (Boltanski and Thévenot, 2006) that is enacted, among other things, through indicators such as the Journal Impact Factor (Rushforth and de Rijcke, 2015; de Rijcke and Müller, 2017). Finn Brunton (this volume, chapter 18) touches upon a similar logic, where he describes how spam and spam journals work off the same socio-technical infrastructures, institutional mechanisms, and rhetoric as "reputable" or "accepted" publishing industries, and hence also fuel the development of these same "legitimate" forms of publishing. The point we make is that publication practices of predatory publishers are being linked to the most important and profitable value systems of the dominant publishing industry and the indicator production market. As such, predatory publishing and its concomitant practices are not outside of the research system but emerge at the heart of them and are embedded within them. These practices in effect drive the existing evaluation logic to the extreme. A crucial question then becomes, *cui bono* (Star, 1995), who actually benefits from this industry?

Generally speaking, there is of course no level playing field in the globalizing system for academic publishing. Arguably, attempts to arrive at such a global, "horizontal" system can in themselves be regarded as a form of vertical domination. On the system level, the publishing industry fortifies boundaries between an "international" West or North on the one hand and a "parochial" East or South on the other. And the case discussed in this chapter shows how predatory publishing can be a vehicle for a particular mode of international, enacted at specific locations in the system. In the Czech Republic and further east, the predatory journals and vanity presses played a role in further empowering skillful local researchers who used the new industry to boost their publication records, international visibility, and the financial status of their institutions (for instance by gaining RIV points for books published by international "vanity" presses[13]). The, at first sight, useful term "predatory publishing" or "predatory journals" may be largely misleading, because it obscures much the agency of individual actors in using these outlets to their advantage. In the case at hand, scholars were hardly "prey," as they found clever ways of gaming the assessment system.

The Czech case makes clear how the predatory publishing industry thrives mainly by being successfully parasitic on existing forms of conduct and material infrastructures for publishing and evaluating research—without fully incorporating its quality control mechanisms (including absence of "proper" peer review and fake editorial boards). But this lack of explicit quality control procedures should not be overemphasized. Some of them apparently have some quality control, and rather than belonging on a blacklist, they operate in a gray zone—into which some established quality journals may now be falling as well with the increased global pressures on production and auditable performance, which deprives the publication system of available competent reviewers and editors. We think the *excessive parasitism* of the "predatory" journals is much more crucial. Many of them deliberately operate on the edges of dominant publication and citation infrastructures, hosted by big commercial publishers. A lot of these journals originate from the "East," and these journals permeate the "global" publishing industry when they are indexed in the WoS and—particularly—SCOPUS.[14] The latter's reputation is based on being the "largest abstract and citation database of peer-reviewed literature," providing a "comprehensive overview of the world's research output."[15] This is obviously a rather problematic statement when the company cannot in practice control this international certification, and is nonetheless taken as proxy for quality in many evaluation systems.

Although the critique of predatory publishing does indeed lead to some sanitization efforts (codes of conduct, blacklists, and whitelists), thus far it has not triggered any serious kind of more radical reform of the publishing and evaluation infrastructure. This may partly be because it is too soon. It could also be due to the fact that purification and policing efforts are often based on the ideal of a unified science system, with internationally shared views "from nowhere" about what constitutes "bad" and "proper" scientific conduct. Such an ideal is doomed to fail when we see how different actors within science systems create and re-enforce distinctive normative hierarchies between the international, the national, and the local: journals, databases, evaluators, consultants, publishers, and also researchers. Some assessment systems are in fact beginning to recognize the need for contextual evaluation (in terms of disciplines and fields) and the complex relation between the international, national, and local. But there still is a long way to go before the research policy and wider academic communities acknowledge that the more, the faster, and the more international need not always be the better.

Notes

1. Work of Sarah de Rijcke on the chapter was supported by the European Union–funded H2020 Project PRINTEGER and by the Technische Universität München-Institute for Advanced Study, funded by the German Excellence Initiative. Work of Tereza Stöckelová on the chapter was supported by grant no. 15–16452S, awarded by the Czech Science Foundation.

2. For details, see the website of the Government Office for Science, Research, and Innovations: http://www.vyzkum.cz/FrontClanek.aspx?idsekce=1028.

3. For example, for papers in WoS journals, the value would be counted on the basis of the position of the journal in disciplinary ranking in WoS but it would include other parameters set up in the evaluation methodology. For patents, the value would depend on whether it is a EU, US, or Japanese patent (one hundred points), a Czech or other national patent (fifty points), or other patent (twenty-five points) (Office of the Government of the Czech Republic, 2013). For a detailed discussion of the evaluation system and its evolution, see Linková and Stöckelová (2012), Stöckelová (2012), Good et al. (2015), and Miholová and Majer (2016).

4. See the archived version of his personal web page: http://web.archive.org/web/20150514044622/http:/www.strielkowski.com/publications (accessed March 30, 2017).

5. An example is the *International Economics Letters*, where he serves as an editor in chief: http://www.ielonline.eu/journal-archive/volume-2-issue-4–2013-/ and http://www.ielonline.eu/editorial-board (accessed May 12, 2016). Another example is the SCOPUS indexed journal *Economics and Sociology*, where he is listed as a board member with the title of "associate professor"—http://www.economics-sociology.eu/?eneditorial-board,3 (accessed May 12, 2016)—a title that incidentally does not match with the information he provides on his own website, according to which he became an associated professor only in 2017 (at the Moscow Power Engineering Institute, Russia) (http://www.strielkowski.com/bio [accessed May 12, 2016]). The researcher also publishes a journal *European Review of Social Sciences* himself through his limited liability company Univerzitní Servis—http://rejstrik.penize.cz/29000335-univerzitni-servis-s-r-o (accessed May 13, 2016), a journal of which he is also an editor: http://erss2.webnode.cz/editors (accessed May 13, 2016).

6. The famous "Beall's list" of predatory publishers and stand-alone predatory journals was created and maintained in the period of 2012 to 2016 by the University of Colorado, Denver, librarian Jeffrey Beall at https://scholarlyoa.com. The list was unexpectedly shut down in January 2017 (Silver, 2017). Refusing at first to comment on the reasons, Beall later stated: "In January 2017, facing intense pressure from my employer, the University of Colorado, Denver, and fearing for my job, I shut down the blog and removed all its content from the blog platform" (Beall, 2017).

7. A summary of the case in English from the whistle-blowers is available at https://zaetickepublikace.wordpress.com/2015/11/17/facts-about-the-critique-of-questionable-publishing-practices-at-the-institute-of-communication-studies-and-journalism-faculty-of-social-sciences-at-charles-university-prague (accessed May 12, 2016).

8. It is noteworthy that his operation looks similar to various hoaxes testing the system, which are described in the fourth section of this volume. However, it was not revealed by the author but his department colleagues. Only then did he call it an "academic joke," adding that "many academics enjoy playing similar jokes" (see http://zaetickepublikace.webnode.cz/questionable-publishing-practices-or-questionable-academics-a-story-from-the-faculty-of-social-sciences-charles-university-in-prague [accessed June 14, 2017]). His newest joke than may be his letter sent to and published in *Nature* in April 2017 in which he praises the Beall's list of predatory publishers and calls for ethics committees to "draw up guidelines for distinguishing reputable from disreputable journals" (https://www.nature.com/nature/journal/v544/n7651/full/544416b.html [accessed June 14, 2017]).

9. http://www.cuni.cz/UK-5554.html (accessed May 12, 2016).

10. http://antipredator.vedazije.cz/index.php?action=switchvalidrecords&show=institution&name=8029 (accessed May 12, 2016).

11. In the Czech Republic, that is. He did find a job at the University of Loughborough in the United Kingdom, and so did the perpetrator, who later became, for some time, a research associate at Cambridge University's Energy Policy Research Group (http://www.strielkowski.com/bio [accessed June 4, 2017]).

12. A petition in support of the whistle-blower was signed by more than one hundred academics—see https://zaetickepublikace.wordpress.com/2015/12/02/prohlaseni-za-publikacni-etiku-a-svobodu-kritiky-v-socialnich-vedach (accessed May 12, 2016). Only his limited individual research grant funding from an external agency would continue, but not the institutional funding he received up to this point.

13. http://antipredator.vedazije.cz/index.php?show=institution&name=8029 (accessed May 14, 2016).

14. See the study by Macháček and Srholec (2017) documenting the sharp rise in recent years of the number of predatory journals identified according to Beall's list in Scopus, with authors of the paper primarily based in the middle-income countries of Asia and North Africa.

15. https://www.elsevier.com/solutions/scopus (accessed May 15, 2016).

References

Beall, Jeffrey. 2012. "Predatory Publishers Are Corrupting Open Access." *Nature* 489:179. doi:10.1038/489179a.

Beall, Jeffrey. 2017. "What I Learned from Predatory Publishers." *Biochemia Medica* 27(2):273–279. https://doi.org/10.11613/BM.2017.029.

Boltanski, Luc, and Laurent Thévenot. 2006 [1991]. *On Justification. The Economies of Worth*. Princeton, NJ: Princeton University Press.

de Rijcke, Sarah, and Ruth Müller. 2017. "Thinking with Indicators. Exploring the Epistemic Impacts of Academic Performance Indicators in the Life Sciences." *Research Evaluation*. https://doi.org/10.1093/reseval/rvx023.

Good, Barbara, Niki Vermeulen, Brigitte Tiefenthaler, and Erik Arnold. 2015. "Counting Quality? The Czech Performance-Based Research Funding System." *Research Evaluation* 24(2):91–105. https://doi.org/10.1093/reseval/rvu035.

Lin, Wen-Yuan, and John Law. 2013. "Making Things Differently: On 'Modes of International.'" *CRESC Working Paper No. 129*. Centre for Research on Socio-Cultural Change (CRESC), Faculty of Social Sciences: The Open University. http://www.cresc.ac.uk/publications/making-things-differently-on-modes-of-international.

Lin, Wen-Yuan, and John Law. 2014. "A Correlative STS: Lessons from a Chinese Medical Practice." *Social Studies of Science* 44(6):801–824.

Linková, Marcela, and Tereza Stöckelová. 2012. "Public Accountability and the Politicization of Science: The Peculiar Journey of Czech Research Assessment." *Science and Public Policy* 39(5):618–629.

Macháček, Vít, and Martin Srholec. 2017. “Predatory Journals in Scopus.” *Study 2/2017, IDEA—CERGE-EI*, Praha. https://idea-en.cerge-ei.cz/files/IDEA_Study_2_2017_Predatory_journals_in_Scopus/mobile/index.html.

Miholová, Katerina, and Ondrej Majer. 2016. *Analýza Mechanismu Metodiky Hodnocení Výsledků [An Analysis of the Mechanism of the Evaluation Methodology]*. Government of the Czech Republic, Section for Science, Research, and Innovation. Unpublished report.

Office of the Government of the Czech Republic. 2013. *Methodology of Evaluation of Research Organizations and Evaluation of Finished Programmes (Valid for Years 2013—2015)*. http://www.vyzkum.cz/storage/att/A7FE6F4477F5064B57B683C62C4A4CFD/Methodology.pdf.

Rushforth, Alexander D., and Sarah de Rijcke. 2015. “Accounting for Impact. The Journal Impact Factor and the Making of Biomedical Research in the Netherlands.” *Minerva* 53:117–139.

Silver, Andrew. 2017. “Controversial Website That Lists ‘Predatory’ Publishers Shuts Down Librarian Jeffrey Beall Won’t Say Why He Has Unpublished His Widely Read Blog.” *Nature*, 18 January. http://www.nature.com/news/controversial-website-that-lists-predatory-publishers-shuts-down-1.21328. doi:10.1038/nature.2017.21328.

Star, Susan Leigh. 1995. “The Politics of Formal Representations: Wizards, Gurus, and Organizational Complexity.” In Susan Leigh Star (ed.) *Ecologies of Knowledge. Work and Politics in Science and Technology*. Albany, NY: State University of New York Press.

Stöckelová, Tereza. 2012. “Immutable Mobiles Derailed: STS and the Epistemic Geopolitics of Research Assessment.” *Science, Technology, and Human Values* 37(2):286–311.

8

Pressures to Publish: What Effects Do We See?

Daniele Fanelli

Concerns for the negative effects of pressures to publish date back at least to the 1950s (Siegel and Baveye, 2010; see also Alex Csiszar in this volume, chapter 1) and today are more widespread than ever. There is virtually no contemporary article that, in analyzing or commenting on issues of research integrity, will fail to suggest that scientists might be increasingly engaging in problematic research practices. At the very least, it is typically argued that scientists may be cynically "salami-slicing" their results (i.e., fractioning them to maximize publication output), but multiple other detrimental practices, right up to the most egregious scientific crime of data fabrication, are suggested to represent plausible strategies to "game" a system that imposes increasingly unreasonable productivity expectations (National Academies of Sciences, Engineering, and Medicine, 2017). Such "pressures to publish" might be imposed upon scientists explicitly by their employing institutions but also implicitly, through the institutional use of faulty metrics of publication quantity and quality (figure 8.1).

This narrative is logically consistent and plausible, but is it correct? The empirical evidence that is most typically invoked in support of the pressures to publish hypothesis comes from anonymous surveys and qualitative studies that observed a connection between reported pressures to publish and likelihood to observe or indulge in questionable behaviors (De Vries, Anderson, and Martinson, 2006; Davis, Riske-Morris, and Diaz, 2007; van Dalen and Henkens, 2012; Tijdink, Vergouwen, and Smulders, 2013). This evidence, however, has clear limitations. Surveys can valuably inform us about what researchers believe, what they say, and what they think they have experienced, but surveys do not necessarily tell us what actually occurs in the general population of scientists. Moreover, results of surveys and interviews are not easy to compare across studies, making it hard to verify whether the problem of pressures to publish has worsened over time as suggested.

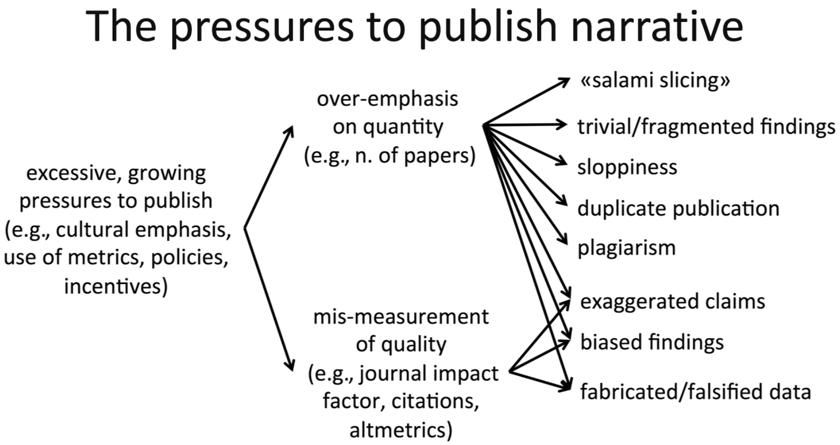

The pressures to publish narrative
excessive, growing pressures to publish (e.g., cultural emphasis, use of metrics, policies, incentives)
over-emphasis on quantity (e.g., n. of papers)
mis-measurement of quality (e.g., journal impact factor, citations, altmetrics)
«salami slicing»
trivial/fragmented findings
sloppiness
duplicate publication
plagiarism
exaggerated claims
biased findings
fabricated/falsified data

Figure 8.1
Schematic summary of hypothesized causal links between pressures to publish and research malpractice, as widely discussed in the literature.

A second and seemingly more direct source of evidence about scientific misconduct appears to be provided by data on retractions of scientific papers. Retractions are mostly the consequence of scientific misconduct (Fang, Steen, and Casadevall, 2012), and the fact that they are more frequent in high-impact journals and that their total number has grown over the years seems to confirm the worse predictions of the pressures to publish narrative (Fang and Casadevall, 2011). Such interpretations, however, are demonstrably premature and likely incorrect. Retractions are an editorial tool of recent invention, and the number of retractions issued per year reflects primarily, if not entirely, the growth in the number of journals that have retraction policies. Back in 2004, only twenty-one percent of high-impact medical journals had a policy to retract papers, whereas in 2014, the percentage had increased to sixty-five percent (Resnik, Wager, and Kissling, 2015). This datum illustrates, on the one hand, how much progress has been made in setting up a system of retractions and yet, on the other hand, how far the system still is from operating at full regime. It is easy to show that the number of retractions no longer appears to be increasing if it is adjusted by the number of journals that are actually issuing retractions, indicating, in other words, that the number of retractions per retracting journal has remained stable for decades (Fanelli, 2013). Moreover, high-impact journals were the first and most proactive adopters of retraction policies (Resnik, Peddada, and Brunson, 2009). This fact alone can explain why high-impact journals have higher retraction rates, even ignoring additional factors like the higher level of scrutiny that these journals are subject to (see chapter by Ivan Oransky, this volume, chapter 10). Therefore, patterns characterizing the prevalence of retractions are not a valid indicator of a possible growing problem with pressures to publish.

The most solid—but still observational and indirect—evidence of a growing problem with pressures to publish comes from statistical analyses of the literature. At least three independent studies suggest that, over the last few decades, abstracts of scientific papers have reported increasingly positive or statistically significant results (Pautasso, 2010; Fanelli, 2012a, 2014; de Winter and Dodou, 2014). Furthermore, at least one study noticed that positive results might be more likely to be reported by abstract of papers from academically productive areas in the United States (Fanelli, 2010a), and at least four meta-meta-analyses in the social and behavioral sciences observed that academically productive countries, and particularly the United States, might publish findings that systematically overestimate underlying effects (Doucouliagos, Laroche, and Stanley, 2005; Munafo, Attwood, and Flint, 2008; Fanelli and Ioannidis, 2013; Fanelli, Costas, and Ioannidis, 2017).

Even meta-analytical evidence, however, offers no conclusive proof of a negative effect of pressures to publish. By drawing correlations at the national level, all the studies listed above are at risk of "ecological fallacy," because correlations observed at the national level might not reflect correlations occurring at the individual level. Moreover, higher rates of positive, statistically significant, and/or extreme results could be produced by mechanisms that have little connection to conscious biases, let alone scientific misconduct (see further discussions in Fanelli [2010a, 2010b]).

Two Empirical Blows to the Pressures-to-Publish Narrative

The most direct assessment of the pressures-to-publish narrative (figure 8.1) comes, to the best of my knowledge, from a series of studies that I recently conducted with several colleagues, which tested multiple hypotheses about determinants of research integrity, misconduct, or bias. The earliest such study, a collaboration with Rodrigo Costas from Leiden University and Vincent Larivière from University of Montréal, examined a large sample of retractions and corrections issued in the years 2010 and 2011 (Fanelli, Costas, and Larivière, 2015). We retrieved the original papers to which the retraction and correction notes referred, and for each of these papers, we retrieved two matched controls—that is, papers that had been published in the same journal and issue but that had not been corrected or retracted. For each of these papers (a total sample of 611 retracted papers, 2,226 corrected papers, and 5,466 controls), we reconstructed the publication profile of first and last authors and recorded other characteristics of study and authors that common hypotheses made in the literature would predict to represent risk factors for scientific misconduct. We predicted that pressures to publish and other risk factors for scientific misconduct should increase the likelihood of retractions and be neutral or decrease the likelihood of corrections, because retractions are usually the consequence of scientific misconduct (Fang, Steen, and Casadevall, 2012), whilst corrections are usually spontaneously solicited by authors (Fanelli and Ioannidis, 2013).

Results showed that the likelihood to be the author of a corrected paper was similar across countries, whereas that of a retracted paper varied substantially. The variance in a country's risk of retraction was partially explained by national publication incentive policies, but not in the direction predicted by the pressures-to-publish narrative. Countries in which high-impact publications are rewarded with cash, such as China, Turkey, and Australia, registered the highest risk of retractions. Countries with career-based publication incentives such as the United States, in which pressures to publish are imposed on the individual through the requirements of tenure, showed intermediate levels of risk. Surprisingly, the risk was lowest in countries such as the Netherlands or United Kingdom, in which universities receive public funding in proportion to their ranking in national research assessments. Since in these latter countries researchers have in theory no choice but to comply with their employer's expectations, these are arguably the only countries in which "pressures to publish" are occurring in a literal sense, and in any case represent countries in which pressures are perceived to be highest (van Dalen and Henkens, 2012).

Therefore, our findings were quite different from what the classic pressure-to-publish narrative would have predicted. The countries at greater risk of misconduct appeared to be those in which researchers are not under institutional pressures to publish, but those in which researchers are lured by cash bonuses. If misconduct can be directly ascribed to a cause, that cause seems to be the corruption and greed of individual scientists.

When we looked at the publication profiles of individual authors, these surprising results were confirmed. The most prolific authors, and those who publish in high-ranking journals, were equally or less likely to produce retracted papers, and equally or more likely to author papers that were later corrected—arguably manifesting higher research integrity.

This pattern was visible even when analyses were limited to authors working in the United States.

These findings were remarkably corroborated by two later studies co-authored by Rodrigo Costas, myself, and others. These studies tested exactly the same hypotheses and the same author characteristics as the study described above, but on completely different proxies of research quality and integrity. One study retrieved a total of over three thousand meta-analyses to test if these parameters predicted the likelihood that a study would report overestimated effect sizes, possibly due to research and publication bias (Fanelli, Costas, and Ioannidis, 2017). The other study used a matched-control approach similar to the one used on retracted papers, but this time on papers that had been identified, by direct inspection, as containing image duplications that are likely to result, at least in part, from intentional misconduct (Fanelli et al., 2017). Both of these studies led to very similar conclusions with regards to the pressures to publish hypothesis: the most prolific authors, those who publish in high-ranking journals, and those working in countries in which pressures to publish are supposedly greater were less likely to report exaggerated findings and less likely to publish papers with image duplications. These results are all observational of course, and therefore do not prove that being a prolific author, or working under high pressures to publish, makes you more honest. However, they are clearly completely at odds with a simplistic narrative that associates publishing too much or too ambitiously with being a cheater.

A second blow to the standard pressures-to-publish narrative came when, in collaboration with Vincent Larivière, I assessed whether scientists are actually publishing papers at an increasing rate (Fanelli and Larivière, 2016). In order to do so, we tracked the individual publication profiles of researchers in all disciplines throughout the twentieth century. From an initial sample of over 540,000 individual authors that we could identify with relative accuracy, we selected those whose main affiliation was in North America, Europe, or Australia/New Zealand, and further limited the sample to authors who had published at least two papers over a period of fifteen years following their first recorded publication (collecting a total sample of 41,000 authors who had co-authored over 760,000 papers between the years 1900 and 2013). Our analysis focused on the first fifteen years of research productivity, because this is an early-career phase in which pressures to publish are presumably highest.

A superficial look at the total number of papers ascribed to each individual author would support the perception that scientists are publishing more. The total number of papers associated with an author's name has increased in most disciplines, including in recent decades, a period in which pressures to publish have arguably become more intense and for which our data is likely to be more accurate. However, the average number of co-authors of these papers had increased as well, and at an accelerating rate. This factor cannot be ignored when estimating scientists' net publication rate.

When we counted publications fractionally, by dividing scientists' total number of publications by their average number of co-authors, the resulting publication rates show no marked increase, and were actually flat or declining in most disciplines. One could argue that co-authorship criteria have simply changed, and that not all names in a manuscript have contributed an equal amount of effort to the publication. However, we also limited the analysis to papers in which our sample of researchers had appeared as first author, a position that in most disciplines identifies the team member who mostly contributed to the publication. Again, we observed no increase. These trends occurred similarly across countries, and multiple secondary analyses suggest that these results are not only robust, but actually rather conservative. For example, when we extended the career time window to twenty-five years, results were very similar, whereas when we restricted it to eight years, publication rates were significantly declining for most disciplines (all robustness results and primary data are provided in the supplementary information of Fanelli and Larivière [2016]).

It must be emphasized that the conclusions of all these large-scale, quantitative studies apply to the average trend. As one should expect, our sample included cases of extremely productive individuals (e.g., researchers who managed to co-author hundreds of papers in just a few years). The number of these cases is likely to have increased during the century, if anything because the total number of scientists has increased. These highly productive individuals are, almost by definition, likely to be widely known, and their higher visibility might reinforce the perception that scientific productivity has risen to excessive levels. However, these extremely prolific authors are amply counterbalanced, at the other extreme, by individuals who publish few papers and yet seemingly do not drop out of a scientific career. On average, therefore, the publication rate of scientists has not increased.

An effect of overexposure similar to the one described above might explain why data fabrication tends to be associated with hyperproductivity. In the aforementioned study on retractions, we noticed that names associated with multiple retractions tend to be, not too surprisingly, highly productive authors (Fanelli, Costas, and Larivière, 2015). The more retractions a case of misconduct brings about, the more exposure it will get in the media (not in small part thanks to the work of Ivan Oransky, this volume, chapter 10), and this might reinforce the public perception that scientific misconduct and unrealistic publication performance are connected.

In sum, according to our findings, scientists today are not publishing, on an individual basis, at higher rates than their colleagues in the 1950s. Today's scientific CVs do list more papers than they used to, but this occurs primarily because scientists today collaborate much more, sharing their efforts as well as their publications. Our results confirmed, with a more rigorous analysis of individual publication patterns, what a simple comparison of the yearly numbers of papers and authors had suggested long ago (de Solla Price, 1980).

Salami-Slicing Collaborations and Sandwiching Results as Alternative Gaming Strategies

In the past, I have published evidence that I interpreted as supporting the pressures-to-publish hypothesis (see Fanelli 2010a, 2012a). In light of these new findings, I am keen to revise my beliefs. Do I believe that pressures to publish are a myth and that we have nothing to worry about? Not at all. The rise of metrics-based performance evaluation is a historical fact, and if scientists claim to feel pressured to publish, there is no reason to doubt them. If these facts didn't suffice, my own personal experience leads me to believe that important and questionable changes are occurring in scientific practices, and that publication performance evaluation has much to do with these changes.

However, I believe that the popular narrative that links pressures to publish to a growing problem of misconduct might be incorrect. Upon closer scrutiny, the direct connection this narrative draws between pressures to publish and misconduct is too simplistic, and the claim that scientists today are publishing at increasing rates is not supported by evidence. It is therefore a double mistake to combine the two claims and conclude that scientific misconduct and other questionable research practices are becoming more prevalent because of growing pressures to publish.

If overproductivity is not directly distorting contemporary science, then how can we explain the aforementioned evidence that null and negative results are proportionally decreasing and especially so in scientifically "productive" countries? I will suggest two phenomena, neglected by past analyses, that might mediate the connection between pressures to publish and questionable research practices.

The first phenomenon is the growth in collaboration size, a trend amply documented in all disciplines and confirmed by my own research, as discussed above. Scientists are likely to have increased their collaboration rates in response to pressures to publish, following a strategy that was openly recommended in the literature (for example, Hayer et al., 2013). Indeed, the growing complexity of research appears to be insufficient to justify, alone, the recent rise in co-authorship, at least in biomedical research (Papatheodorou, Trikalinos, and Ioannidis, 2008).

Larger collaborations might run a higher risk of bias and misconduct. Collaborations represent an investment of resources and professional reputation of considerable size; their members, therefore, might be under comparably high pressures to make such investment "pay off." We can therefore hypothesize that the design and conduct of collaborative studies might be exceedingly oriented toward producing "publishable" results, possibly at the expense of scientific rigor and integrity. Errors, bias, and misconduct, moreover, might be more likely to escape scrutiny in collaborations from different fields and institutions, because cultural and geographic distances impede mutual criticism and supervision. Preliminary data support this hypothesis, by suggesting that long-distance collaborations are associated with higher rates of positive results (Fanelli, 2012b) and do not protect against the risk of retractions (Fanelli, Costas, and Larivière, 2015; Fanelli, Costas, and Ioannidis, 2017; Fanelli et al., 2017).

Collaborations might be proliferating excessively because the h-index and other popular metrics of research productivity and impact are not adjusted for co-authorship or actual contribution. From a technical point of view, this is a highly questionable choice, which presumably was made not for sound mathematical reasons, but to keep publication metrics simple and to encourage collaboration, which is assumed to be a positive force in science. Paradoxically, however, by not imposing any costs to co-authorship, current metrics might discourage real cooperation, because higher "performance" scores can be accrued by scientists who fragment their effort into as many collaborations as possible and contribute the minimum possible to each (Kaushal and Jeschke, 2013). Scientists, I am suggesting, might be increasingly "salami-slicing" not their results, but their collaboration efforts.

The second phenomenon is a general increase in the length and information content of scientific papers. Such a trend has been documented in biomedical research (Vale, 2015) and was supported in our cross-disciplinary data, which showed that the average page length of papers has increased throughout the twentieth century (Fanelli and Larivière, unpublished observation). This phenomenon, which openly contradicts the hypothesis that scientists are increasingly "salami-slicing" their results, is likely to reflect the growing complexity of scientific research and the availability of ever more powerful computational tools. Pressures to publish might reinforce this trend, by compelling scientists to "pad" their papers with unnecessarily large numbers of data points and secondary analyses, in order to boost their papers' chances of acceptance by the journal. Sheer quantity of data and analyses might be increasingly used as a cheap substitute for quality (that is, for convincing conclusions drawn from well-researched studies and cleverly designed experiments). Moreover, pressures to publish might prompt scientists to overpromote their work by emphasizing, simplifying, and perhaps exaggerating the strength and originality of their findings. Under this scenario, negative and nonstatistically significant results might not be, as usually suggested, left lying in the proverbial file drawers but might be buried instead into longer and more complex papers, which in titles and abstracts highlight only positive findings. The salami slices, in other words, are not taken apart and published individually, but used to pad large and seemingly rich panini, only the juiciest fillings of which are allowed to stick out and be exposed in titles and abstracts. Rather than "salami-slicing" and selecting results, I suggest, scientists may be increasingly "sandwiching" them.

In conclusion, contrary to what is commonly argued in the literature, there is little direct evidence to suggest that pressures to publish have increasingly undermined the integrity of scientists. However, it is reasonable to hypothesize that such pressures might have contributed to alter scientists' publication and collaboration practices. Scientists might be gaming metrics by "sandwiching" their negative results and "salami-slicing" not their publications, as commonly believed, but their collaborations. These strategies might have been effective to the point of going largely unnoticed by generations of scholars concerned by the negative effects of pressures to publish.

References

Davis, M. S., M. Riske-Morris, and S. R. Diaz. 2007. "Causal Factors Implicated in Research Misconduct: Evidence from ORI Case Files." *Science and Engineering Ethics* 13:395–414. doi:10.1007/s11948-007-9045-2.

de Solla Price, D. J. 1980. “The Citation Cycle.” In *Key Papers in Information Science*, edited by B. C. Griffith, 195–210. White Plains, NY: Knowledge Industry Publications.

De Vries, R., M. S. Anderson, and B. C. Martinson. 2006. “Normal Misbehavior: Scientists Talk About The Ethics Of Research.” *Journal of Empirical Research on Human Research Ethics* 1(1):43–50.

de Winter, J., and D. Dodou. 2014. “A Surge of P-Values Between 0.040 and 0.049 in Recent Decades (But Negative Results Are Increasing Rapidly Too).” *PeerJ* PrePrints (2):e447v3.

Doucouliagos, H., P. Laroche, and T. D. Stanley. 2005. “Publication Bias in Union-Productivity Research?” *Relations Industrielles-Industrial Relations* 60(2): 320–347.

Fanelli, Daniele. 2010a. “Do Pressures to Publish Increase Scientists’ Bias? An Empirical Support from US States Data.” *PLoS One* 5(4). doi:10.1371/journal.pone.0010271.

Fanelli, Daniele. 2010b. “‘Positive’ Results Increase Down the Hierarchy of the Sciences.” *PLoS One* 5(3). doi:10.1371/journal.pone.0010068.

Fanelli, Daniele. 2012a. “Negative Results Are Disappearing from Most Disciplines and Countries.” *Scientometrics* 90(3):891–904. doi:10.1007/s11192-011-0494-7.

Fanelli, Daniele. 2012b. “When East Meets West…Does Bias Increase? A Preliminary Study on South Korea, United States and Other Countries.” Proceedings of Thirteenth COLLNET Meeting, Seoul, South Korea.

Fanelli, Daniele. 2013. “Why Growing Retractions Are (Mostly) a Good Sign.” *PLoS Med* 10(12):e1001563. doi:10.1371/journal.pmed.1001563.

Fanelli, Daniele. 2014. “Reply, to de Winter and Dodou (2014): Growing Bias and the Hierarchy Are Actually Supported, Despite Different Design, Errors, and Disconfirmation Biases.” http://danielefanelli.com/files/reply to de Winter & Dodou 2014.pdf.

Fanelli, Daniele, and John P. A. Ioannidis. 2013. “US Studies May Overestimate Effect Sizes in Softer Research.” *Proceedings of the National Academy of Sciences of the United States of America* 110(37):15031–15036. doi:10.1073/pnas.1302997110.

Fanelli, Daniele, Rodrigo Costas, Ferric C. Fang, Arturo Casadevall, and Elisabeth M. Bik. 2018. “Testing Hypotheses on Risk Factors for Scientific Misconduct via Matched-Control Analysis of Papers Containing Problematic Image Duplications.” *Sci Eng Ethics*. doi:10.1007/s11948-018-0023-7.

Fanelli, Daniele, Rodrigo Costas, and John P. A. Ioannidis. 2017. “Meta-Assessment of Bias in Science.” *Proceedings of the National Academy of Sciences* 114(14):3714–3719. doi:10.1073/pnas.1618569114.

Fanelli, Daniele, Rodrigo Costas, and Vincent Larivière. 2015. “Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity.” *PLoS One* 10(6):e0127556. doi:10.1371/journal.pone.0127556.

Fanelli, Daniele, and Vincent Larivière. 2016. “Researchers’ Individual Publication Rate Has Not Increased in a Century.” *PLoS One* 11(3). doi:10.1371/journal.pone.0149504.

Fang, F. C., R. G. Steen, and A. Casadevall. 2012. “Misconduct Accounts for the Majority of Retracted Scientific Publications.” *Proceedings of the National Academy of Sciences of the United States of America* 109(42):17028–17033. doi:10.1073/pnas.1212247109.

Fang, Ferric C., and Arturo Casadevall. 2011. “Retracted Science and the Retraction Index.” *Infection and Immunity* 79(10):3855–3859. doi:10.1128/iai.05661–11.

Grieneisen, Michael L., and Minghua Zhang. 2012. “A Comprehensive Survey of Retracted Articles from the Scholarly Literature.” *PLoS One* 7(10). doi:e4411810.1371/journal.pone.0044118.

Hayer, Cari-Ann, Mark Kaemingk, Jason J. Breeggemann, Daniel Dembkowski, David Deslauriers, and Tobias Rapp. 2013. “Pressures to Publish: Catalysts for the Loss of Scientific Writing Integrity?” *Fisheries* 38(8):352–355. doi:10.1080/03632415.2013.813845.

Kaushal, S. S., and J. M. Jeschke. 2013. “Collegiality Versus Competition: How Metrics Shape Scientific Communities.” *Bioscience* 63(3):155–156. doi:10.1525/bio.2013.63.3.3.

Munafo, Marcus R., Angela S. Attwood, and Jonathan Flint. 2008. “Bias in Genetic Association Studies: Effects of Research Location and Resources.” *Psychological Medicine* 38(8):1213–1214.

National Academies of Sciences, Engineering, and Medicine. 2017. *Fostering Integrity in Research*. Washington, DC: The National Academies Press.

Papatheodorou, S. I., T. A. Trikalinos, and J. P. A. Ioannidis. 2008. “Inflated Numbers of Authors Over Time Have Not Been Just Due to Increasing Research Complexity.” *Journal of Clinical Epidemiology* 61(6):546–551. doi:10.1016/j.jclinepi.2007.07.017.

Pautasso, M. 2010. “Worsening File-Drawer Problem in the Abstracts of Natural, Medical and Social Science Databases.” *Scientometrics* 85(1):193–202. doi:10.1007/s11192-010-0233-5.

Resnik, David B., Elizabeth Wager, and Grace E. Kissling. 2015. “Retraction Policies of Top Scientific Journals Ranked by Impact Factor.” *Journal of the Medical Library Association* 103(3):136–139. doi:10.3163/1536–5050.103.3.006.

Resnik, David B., Shyamal Peddada, and Winnon Brunson, Jr. 2009. “Research Misconduct Policies of Scientific Journals.” *Accountability in Research-Policies and Quality Assurance* 16(5):254–267. doi:10.1080/08989620903190299.

Siegel, Donald, and Philippe Baveye. 2010. “Battling the Paper Glut.” *SCIENCE* 329(5998):1466–1466. doi:10.1126/science.329.5998.1466-a.

Tijdink, J. K., A. C. M. Vergouwen, and Y. M. Smulders. 2013. “Publication Pressure and Burn Out among Dutch Medical Professors: A Nationwide Survey.” *PLoS One* 8(9):6. doi:10.1371/journal.pone.0073381.

Vale, R. D. 2015. "Accelerating Scientific Publication in Biology." *Proceedings of the National Academy of Sciences of the United States of America* 112(44):13439–13446. doi:10.1073/pnas.1511912112.

van Dalen, H. P., and K. Henkens. 2012. "Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey." *Journal of the American Society for Information Science and Technology* 63(7):1282–1293. doi:10.1002/asi.22636.

9

Ghost-Managing and Gaming Pharmaceutical Knowledge

Sergio Sismondo

In his chapter in this volume (chapter 1), and echoed in a number of the other chapters here, Alex Csiszar describes how gaming metrics of science productivity was an immediate, possible consequence of the use of metrics—recognized in Goodhart's law—and an immediate concern in early discussions of metrics derived from the Science Citation Index. Gaming the system occurs when moves that are not against the rules, or that can be made to appear to be not against the rules, lead to some kind of surplus value. The truth of Goodhart's law, then, is a consequence of a more general kind of opportunism that fills "economic" (in a broad sense) niches. Particular instances of gaming can reveal both some possible opportunistic actions and the economic structures that make them possible. In this chapter, I focus more on the former, but try not to lose sight of the latter.

Here I describe an arena where related economies meet, and where various goods can be created by moving resources from one into another. The economies in question are those of medical science, medical practice, and pharmaceutical marketing. One of the results of the meeting is the publication of medical journal articles that look like reports of academic science, but that have been largely or wholly created by many corporate actors working together, with the ultimate goal of influencing prescriptions. Pharmaceutical companies and their agents control or shape, in ways that are not entirely visible, multiple steps in the research, analysis, writing, publication, and dissemination of significant amounts of medical science. I call this the "ghost management" of medical science.

Most of the clinical trial research that the pharmaceutical industry funds is handled by contract research organizations (CROs), companies that can run all different aspects of clinical trials. The data that CROs produce is typically analyzed by pharmaceutical company statisticians, reported in articles written by medical writers, and guided through to publication by dedicated publication planners. Publication plans parcel data and other information for journals in ways that will be recognized and respected by various important physician and researcher audiences. Typically, it is only after the articles are drafted that authors are recruited. Authors are generally seen as "key opinion leaders" or "KOLs": doctors and researchers valued for their status within—or at least participation in—their specialties, and with whom the companies have established relations. In addition to having credibility, authors may be chosen because they already agree with the conclusions of the articles, or because they are prospective speakers on the drugs at issue. The articles form the basis of company-funded presentations by KOLs at conferences, in continuing medical education courses, and in innumerable small-scale events in clinics and restaurants, as part of what are known as "speaker programs." Company sales representatives and medical science liaisons—the latter are staff who interact with physicians and researchers primarily about the science that supports products—also distribute articles in visits to and exchanges with physicians.

Even in the sketch in the previous paragraph, we can see the pharmaceutical industry leveraging academic value by gaming academic communication. Pharmaceutical companies have joined the communication structures of academic science, making contributions that look as much as possible like good academic science, but that help to support commercial aims, in the form of encouragement to physicians to prescribe. In their speaker programs and door-to-door delivery of information, the companies have added new forms of communication, built on top of traditional academic forms. The medical science that pharmaceutical companies produce and circulate leads to drug prescriptions. In the following, I add some details to the picture I have just presented, displaying different forms of the leveraging of academic value, including one that takes us into the realm of metrics.

My research here draws on a wide variety of kinds of communications internal to the industry. In particular, it is largely based on my and two research associates' attendance at a number of industry conferences between 2007 and 2017, where people who are insiders make presentations on issues of publication planning, KOL management, or speaker programs (for more details see Sismondo and Chloubova, 2016; Sismondo, 2018). Much of my work focuses on what Finn Brunton (this volume, chapter 18) calls "secondary markets" around pharmaceutical research and marketing. For the sake of prudence and in accordance with my research protocol, I anonymize all sources, even though some of them were speaking in essentially public venues.

Publication Planning

The construction of industry-sponsored articles almost always involves publication planners. According to planners, their work can and should start even before the research does, contributing to research design, mapping out key messages, and designating different articles for different audiences and journals. Once the research is available, publication planners hire medical writers, contact potential authors, negotiate with various interests and departments within the pharmaceutical companies, and shepherd the articles through journals' submission and revision procedures (Auti et al., 2016). Ms. I, a planner working within a pharmaceutical company, says:

> This is what utopia looks like from an industry perspective. We have agreement and alignment on a plan, not even just a publication, a full plan, investigators on board, agencies lined up, everybody ready to play and we're going to get this done in a timely way, in an orderly fashion, and things work like clockwork.

The best publication plans comprehensively address research, development, presentations, and publications; appendices give the relevant data for each of the meetings and journals to which abstracts and papers will be submitted—the audiences they reach, their impact factors, their rejection rates, and publication lead times (Complete Healthcare Communication, 2006). A plan may also describe other communication opportunities, such as symposia and roundtables, journal supplements, advisory board meetings, monographs, slide programs, formulary kits, and more. Planners have been known even to create entirely new journals for particular projects, though that seems to be a rare and scandalous occurrence (Grant, 2009).

At the same time, planners should be responsive to changing circumstances and to the changing priorities of the company. Thus they might need to arrange for the production of a letter to the editor in response to an unfavorable study or the needs of a public relations campaign. Mr. D, a planner working for a large pharmaceutical company, illustrates this when he talks about supporting his company's key messages:

> At the beginning of the year, we kind of have a scientific strategy for every product, saying, y'know, these are the key messages that we're hoping to get out, depending on what clinical data we have available. We'll look at all the points that the upper management folks would like us to try and see if we have the data to address, and then we'll go through it point by point and try to see.

The biggest growth of publication planning occurred in the 1990s and probably was connected to other changes in the global pharmaceutical economy. That decade saw an enormous increase in global sales of pharmaceuticals, at an average rate of over ten percent per year (World Health Organization, 2004). This surge was spurred by an increasing number of blockbuster drugs and consistent high sales growth in the United States. There was also a change in the structure of research: in 1990, seventy percent of pharmaceutical industry research funding went to universities and teaching hospitals, whereas in 2000, seventy percent went to CROs (Mirowski and Van Horn, 2005), a level that now seems stable (Westrock, 2016). The simultaneous rise of the publication planning and CRO industries almost certainly is not coincidental, since CRO research can be planned and harnessed to marketing goals more easily than can academic research: CROs, unlike academics, have little interest in publishing the results of their studies.

Planners recognize that their work has marketing value, and publication planning agencies often advertise their work in terms of the contribution it can make to marketing. Tongue in cheek, industry consultant Ms. S asks the audience at one meeting, “By the way, is anything you do ever used in a promotional context? Oh yeah!” The promotion can be very broad. A publication plan accompanying the launch of a likely blockbuster drug can include more than fifty articles published over three or so years (Healy and Cattell, 2003; Ross et al., 2008). A chart presented at an industry seminar entitled “Publication Planning 101,” showing the number and type of publications per year for a fictional new product, displays roughly ninety articles to be published over the course of five years. These are labeled: clinical efficacy, clinical pharmacology, review, case report, letter to the editor, and quality of life. In another context, Ms. S says, “The newest thing right now is disease states.… You all know what I’m talking about, where you don’t mention the name of the drug but you talk about the disease.”

Pharmaceutical companies typically arrange for KOLs to serve as all or the majority of authors on manuscripts. By using KOLs as authors, publication planners can give articles a veneer of having been written by independent researchers, instead of by a coordinated industry team. A KOL author thus increases the perceived credibility of an article and also hides features of the research process and analysis. Because of pressure from journal editors, the work of medical writers is increasingly being recognized in the acknowledgments sections of articles, but company statisticians and researchers, reviewers from an array of departments, and publication planners are rarely mentioned.

In general, KOL authors are very unlikely to have worked closely with the data they are reporting. Pharmaceutical companies initiate and fund the planning, research, analysis, writing, and placing of articles, and typically maintain control of data throughout. Industry representative Dr. Q even argues that authors should not be given *access* to the data, because they may lack skill, and they may have their own agendas: "As the owners of the study database, the sponsors will decide who will have access to the database.… PhRMA [Pharmaceutical Research and Manufacturers of America, the US industry lobby group] companies commit to making a summary of the results available to the investigators." According to Mr. B, working for an independent planning agency, fifty percent of companies show only the penultimate draft to authors, to solicit their input. As a result, ghost-managed articles almost always violate naïve readers' expectations about their trajectories, and generally violate journals' authorship criteria—though publication planners try to make it possible to make a case that their authors technically meet those criteria.

The KOLs may have multiple reasons for agreeing to serve as authors on these manuscripts. They add articles to their CVs, and, as I discuss below, these articles are likely to be amply cited. Although pharmaceutical companies do not pay for authorship, they may ask authors to give presentations of, or related to, the research, for which the authors are generously paid. Finally, it can be flattering to be targeted as an expert, and the manuscripts themselves may even contain more flattery, as this short excerpt from a legal deposition of a publication planner, discussing a ghost-managed review article, shows:

> *Q*. All right. So before Dr. M. Brincat [the eventual author] saw the outline, Designwrite [the publication planning firm involved] had done the medical research, the literature research, to determine whether there was sufficient scientific evidence to support a scientific platform for this article. An outline was drafted and then Mr.… approached Brincat and Brincat agreed to be an author; is that correct?
>
> *A*. That is correct, because it mostly cited Dr. Brincat's research. (US District Court, 2006)

Editors of all of the important medical journals are aware of the process, and almost every publication planning conference includes a panel of editors. While the editors typically condemn ghostwriting, they seem to accept that the strong pharmaceutical industry presence in medical research necessitates the ghost management of research and publication. These editors and their journals also value the articles, which, again, tend to be respected and amply cited, and which may turn into immediate revenue if the companies sponsoring them want to buy reprints for distribution. It is striking that, despite the occasional exposé revealing ghostwritten articles, retractions because of industry ghosting are extremely rare or nonexistent (Jones, 2009).

Speaker Programs and Other KOL Activities

For fifty years or more, sales representatives, medical science liaisons, and, recently, independent firms have been identifying potential KOLs, establishing relations with them and developing them into more effective speakers and advocates for companies: “A key task of a pharmaceutical rep’s job is to help transform influential doctors into speakers and consultants who know the rules of the game and are quite adept at negotiating a stipend and ‘working the crowd’” (Oldani, 2004; see also Sismondo, 2018).

The industry recognizes different kinds of KOLs, requiring different forms of interaction. Ordinary physicians—either general practitioners or specialists—are paid to speak to other physicians as members of speaker bureaus for particular drugs: they might address other physicians at lunchtime talks organized by sales reps, or serve as after-dinner speakers at physicians’ events, also organized by sales reps. Medical researchers’ value to pharmaceutical companies might stem from any number of activities: they might be paid to speak to researchers or patient groups or at continuing medical education sessions; they might be consulted on any number of medical, marketing, or research issues; they might serve as authors of ghost-managed medical journal articles; or they might contribute to research either by recruiting patients for trials or by initiating their own trials.

Like publication plans, speaker programs can be large. Pharmaceutical company manager Mr. E, presenting at a KOL management conference, raises the specter of an investigation of a speaker program: “When you say ‘I need seven hundred to one thousand speakers in this activity,’ the questions [that are] going to get pushed back to you in investigations are, ‘Why do you need so many? How many is each speaker going to do? Why did you need a thousand?’” Mr. E’s concern is that investigators will conclude that some speakers are being trained and paid not because they are effective communicators but because they are important prescribers. There is a continuum from KOLs employed primarily to change other physicians’ prescribing patterns to those employed primarily to change their own prescribing patterns (which is generally illegal). The latter suggests a devious way of leveraging academic-like communication structures, influencing people by hiring them to speak.

With physician KOLs, the goals and consequent relationships are straightforward, since the physicians are simply hired to give talks and typically are given zero latitude in their delivery. Researcher KOLs, though, are treated so that they feel more like partners in medical science and education. Interactions with them need to be subtle, especially since much of KOLs' value to the companies stems from their independence from those companies, creating a real tension. Still, the needed independence does not stop KOL management experts from repeatedly indicating that KOLs can be used as important mediators for pharmaceutical companies. Here is Ms. C speaking to an audience of KOL managers and others:

> [A] KOL point person can help you and the organization make sure that you are…identifying the right expert for the right need and able to work with them at the right place and time and be able to deliver a KOL plan that's aligned to their scientific objectives.…Particularly as you start to enter Phase One, Phase Two [trials], and, you know, these molecules are moving along, it looks to have some promise—okay there are unique aspects perhaps about the mechanism of action—it's going to be very important to help start to educate the community, the physician community, the patient community, the professional societies on this mechanism of action [and] on the disease state itself.

Researcher KOLs can smooth the path to acceptance of drugs and diseases by helping to shape the background of accepted issues and opinions in a field. They might participate in industry-sponsored workshops and author key papers, thereby becoming the experts to whom the FDA could turn for advice on drug submissions and to whom the media could turn for interviews and information. In this way, they act as mediators between pharmaceutical companies, the FDA, physicians, and potential consumers (Fishman, 2004).

A Citation Puzzle

The ghost management of medical research presents a citation puzzle linked to companies' gaming of academic communication systems. Gorry (2015) analyzes a group of ninety-two articles known to be ghost-managed, identified in documents from three legal proceedings. Among other things, Gorry notes that ghost-managed articles were cited approximately ten times more often than were typical other articles in the same journals—and almost none of the difference is explained by a difference in prestige of the authors (personal communication). Healy and Cattell

(2003) had earlier analyzed a subset of that group, and, unsurprisingly, the two studies come to some overlapping conclusions. Healy and Cattell compare their group of fifty-five ghost-managed articles on a particular drug with other articles on the same drug published in the same period: ghost-managed articles were cited between 2.4 and 2.9 times more frequently than matched counterparts.

I suggest three possible explanations of the high citation rate of ghost-managed articles, all of which are very likely right, though I can only point to factors that make each one plausible.

First, ghost-managed articles may be more *cite-worthy* than their various counterparts. Pharmaceutical companies sponsor the majority of medium-sized and large clinical trials, the kind of study that is most valued in the medical world. The resources of pharmaceutical companies enable them not only to run solid trials, but also to produce articles that have all the hallmarks of good science. If ghost-managed articles are "counterfeits" and independent ones are "authentic," in this case the counterfeit appears to be as high quality as is the authentic.

Second, ghost-managed articles may be more cited because they have better distribution than their counterparts. As marketing vehicles, these ghost-managed articles need to be read, or at least seen. Pharmaceutical companies have excellent distribution systems for their articles and for the information contained in them. The companies pay for presentations by KOLs at conferences, continuing medical education courses, clinics, and after-dinner events. Sales representatives and medical science liaisons provide reprints of articles to physicians, including academic physicians. Occasionally, companies engage in mass mailings of reprints. Ghost-managed articles, then, are tremendously better circulated than are independent articles.

Third, the ghost management process likely leads to an interesting version of self-citation. A publication plan that involves fifty or a hundred articles provides many potential entries in a reference list. Later articles can cite earlier ones, and all can cite articles from earlier publication plans, and not just earlier articles by particular authors. Describing an episode in her work as a medical writer, Larkin (1999) writes:

> I agreed to do two reviews for a supplement to appear under the names of respected "authors." I was given an outline, references, and a list of drug company-approved phrases. I was asked to sign an agreement stating that I would not disclose anything about the project. I was pressured to rework my drafts to position the product more favorably.

Presumably, the list of references she was given was just as drug company-approved as was the list of phrases—medical writers and publication planners describe the literature review as a key step in the development of an article. Indeed, it would be curious if reference lists were *not* skewed toward the company's previous articles, because those articles would tend to support the company's commercial interests and because those would be the articles or references to hand. If such self-citation exists, it is unusually invisible, in that the self-references are not to an individual, a laboratory, or a department, but rather to a set of publication plans and a company.

Conclusion: Multiple Leverage Points

Pharmaceutical companies have joined academic medicine's research and communication structures. They participate both covertly and overtly, sometimes choosing to build medical knowledge minimally marked by conflicts of interest, and sometimes choosing to establish strong connections between scientific evidence and brands.

Not only do companies participate, they participate with more and better resources than are available to independent academics. Their ability to hire CROs allows them to run randomized, controlled trials worldwide, on drugs for varied conditions; clinical trials produce the kind of knowledge that is most valued within medicine. Companies produce articles for academic journals that look like independently produced articles and have independent academics as authors, but that are likely to be more influential than are independently produced articles.

Not only do companies participate, they innovate. Building on forms like academic conferences and continuing medical education, they have developed sponsored research workshops and speaker programs, at which their KOLs give presentations that have the look and feel of presentations in academic research and education contexts, and sometimes might even be confused for independent work. Along the way, the companies may be using these forms to convince the KOLs themselves to prescribe their products.

The intersection of different economies means that various different goals and metrics are at play here. A ghost-managed article contributes to medical knowledge. For pharmaceutical companies, the content of articles is important, because it can serve as a justification for prescriptions of their products. For those companies, at issue is the monetary return on investment of the publication of research. Although this can be difficult to measure, there are attempts to measure the increased prescriptions resulting from particular articles. For a KOL author, an article represents another line on a CV and a contribution to prestige; it also may lead to industry-paid speaking engagements and to new citations. For a journal, a ghost-managed article contributes to its positive reputation for publishing clinical and other research; the article is also likely to contribute to its impact factor, and if the sponsoring company uses reprints of it for promotion, the article could provide cash revenue.

Consistent with the insights of Marie-Andrée Jacob in this volume (chapter 19), ghost-managed articles are not simple counterfeits or fakes, standing in opposition to authentic or real articles. They are developed and constructed with considerable resources and skill, and rely on rich data. They are widely distributed and have real-world impact in the form of prescriptions. They are often exemplary pieces of medical science—albeit medical science created for marketing purposes, and with commercial interests driving the work. Moreover, these articles rely for their effectiveness on at least a limited amount of collaboration with academic medical science—in the form of offering authorship of articles, with the endorsement that that implies. While, because authors typically do not meet journals' authorship criteria, there is typically misconduct, there is a sense in which the misconduct needs to be carefully teased apart from more prototypical misconduct, such as faking data. Instead, pharmaceutical companies are now the biggest contributors to the evidence base of medicine, ghost-managing apparently high-quality, interest-driven scientific knowledge: gaming academic communication and leveraging academic value for commercial goals.

References

Auti, Prashant, Rishabh Pandey, and Vatsal Shah. 2016. "Project Management in Medical Publication Writing: A Less Explored Avenue in Pharmaceutical Companies and Clinical Research Organisations." *Medical Writing* 25(1):36–44.

Complete Healthcare Communication. 2006. "Complete Healthcare Communication." http://www.chcinc.com/ (accessed December 20, 2006).

Fishman, Jennifer. 2004. "Manufacturing Desire: The Commodification of Female Sexual Dysfunction." *Social Studies of Science* 34(2):187–218.

Gorry, Philippe. 2015. "Medical Literature Imprinting by Pharma Ghost Writing: A Scientometric Evaluation." Paper presented at the Fifteenth International Conference on Scientometrics and Infometrics, Turkey.

Grant, Bob. 2009. "Elsevier Published 6 Fake Journals." *The Scientist.com* (May 7). http://www.thescientist.com/templates/trackable/display/blog.jsp?type=blog&o_url=blog/display/55679&id=55679 (accessed July 19, 2009).

Healy, David, and Dinah Cattell. 2003. "Interface Between Authorship, Industry and Science in the Domain of Therapeutics." *British Journal of Psychiatry* 183:22–27.

Jones, Nicola. 2009. "Analysis of Retractions Puts Spotlight on Academia." *Nature Medicine* 15(10):1101.

Larkin Marilynn. 1999. "Whose Article Is It Anyway?" *Lancet* 354(9173):354:136.

Mirowski, Philip, and Robert Van Horn. 2005. "The Contract Research Organization and the Commercialization of Scientific Research." *Social Studies of Science* 35(3):503–534.

Oldani, Michael J. 2004. "Thick Prescriptions: Toward an Interpretation of Pharmaceutical Sales Practices." *Medical Anthropology Quarterly* 18(3):325–356.

Ross, Joseph S., Kevin P. Hill, David S. Egilman, and Harlan M. Krumholz. 2008. "Guest Authorship and Ghostwriting in Publications Related to Rofecoxib: A Case Study of Industry Documents from Rofecoxib Litigation." *Journal of the American Medical Association* 299(15):1800–1812.

Sismondo, Sergio. 2018. *Ghost-Managed Medicine: Big Pharma's Invisible Hands*. Manchester: Mattering Press.

Sismondo, Sergio, and Zdenka Chloubova. 2016. "'You're Not Just a Paid Monkey Reading Slides'": How Key Opinion Leaders Explain and Justify their Work." *BioSocieties* 11(2):199–219.

US District Court. 2006. "Deposition of Karen D. Mittleman in Re: Prempro Products Liability Litigation." https://www.industrydocumentslibrary.ucsf.edu/drug/docs/#id=sqbw0217 (accessed June 4, 2017).

Westrock. 2016. "Patient-Centric Clinical Trial Solutions: Adherence, Outcomes and Data." https://pages.questexweb.com/WestRock-20160524-Registration.html (accessed May 15, 2016).

World Health Organization. 2004. "The World Medicines Situation." http://www.who.int/medicinedocs/collect/medicinedocs/pdf/s6160e/s6160e.pdf (accessed April 30, 2008).

III

Interventions: Notes from the Field

The detection of new forms of manipulation exceeds the reach, resources, and perhaps even the mandate of traditional governmental agencies like the US Office of Research Integrity. These constraints, however, have been balanced by the emergence of a new generation of collaborative grassroots watchdogs, carnivalesque pranksters, independent organizations, and bloggers. The watchdogs often operate websites featuring public discussion or anonymous tips about publication misconduct that is often propelled by metrics. While watchdogs are often humorous, pranksters use humor and satire as their primary mode of critique to reveal spammish journals and easily gamed metrics. This section offers "notes from the field" by and about some of these new actors—notes that give us a glimpse of how they have emerged, how they operate, and how they have assumed increasingly important roles and credible voices despite a lack of governmental or university affiliations. They also tell us about the criticism, challenges, and legal threats they face.

Ivan Oransky co-founded the Retraction Watch to tell the stories behind article retractions, their typical opacity, and their increasing frequency. Virtually unknown a few decades ago, retractions have become a daily occurrence, their quantity allowing for qualitative and quantitative studies of their patterns. The Retraction Watch website, coupled with its daily email updates, looks at the retractions themselves and what their nature, distribution, and causes can tell us about general, even global, trends in misconduct and metrics manipulations. Brandon Stell's and Boris Barbour's chapter shows that PubPeer—initially fashioned as a web-based journal club—shares many of the motivations that animate Retraction Watch, but focuses on providing a publicly accessible platform for the detailed discussion of scientific articles chosen by the participants. Some questions and doubts are quickly answered by the authors of the articles (who are welcomed to jump into the public discussion), but in other cases, the public questioning unearths more serious problems that may eventually lead to retractions. According to PubPeer, many of these problems are directly relatable to the craze for publishing in high–impact factor journals.

PubPeer has been very successful largely because it ignores metrics and engages in good old reading and critiquing. And while it does not take a position on the claims and allegations put forward by its largely anonymous commentators, these claims and the evidence they are based on is open for everybody to see, even when they may not be too eager to see it. For instance, Catherine Guaspare and Emmanuel Didier document a high-profile case—the "Voinnet affair"—that unfolded largely due to discussions on PubPeer and Retraction Watch. They argue that that case (in which a renowned French scientist lost an academic post and was banned from funding after the discovery of his frequent image manipulations) exemplifies a new "ecology" of misconduct detection in which the new watchdogs, despite having virtually no institutional authority, are able to create such a widespread publicity that forces the authorities to intervene or else appear to condone the misconduct.

Unlike the probing but formally polite discussions that animate PubPeer, Paul Brookes's blog "Science Fraud" chose to deploy snarky language, and to do so from behind the mask of anonymity. This was a strategic choice, hoping that the somewhat "carnivalesque" tone of his commentary would draw additional attention, and thus put pressure on the authorities to investigate its claims. Anonymous and pseudonymous readers emailed him their discoveries of potential misconduct, which he blogged about humorously and sarcastically, while crediting his (masked) tipsters. After six months, however, Brookes's identity was exposed. As a result, he was threatened with lawsuits by those accused of misconduct (a fate that has since been shared by PubPeer)[1] while his university—fearing liabilities—demanded that he choose between the blog and his professorship. The forced termination of "Science Fraud" and the recent legal challenges to PubPeer show how independent watchdogs, thanks to their "crowdsourcing" approach, can be quicker and more effective than traditional "misconduct police" agencies like the Office of Research Integrity. At the same time, their lack of institutional resources and authority, or of a governmental mandate, forces them to develop creative uses of anonymity (for themselves or their sources), which can backfire not only because anonymity can be breached, but also because its use may lead to accusations of opacity and unaccountability. Analogies between the predicament of PubPeer and WikiLeaks are not too far-fetched.

Through a humorous experiment, Cyril Labbé confronts the problem of quantifying academic excellence through ranking scholars according to their citation counts. He created an imaginary persona—“Ike Antkare” (I Can’t Care)—as the author of dozens of computer-generated papers, which he then proceeded to publish to demonstrate how easily Google Scholar’s ranking could be gamed. Thanks to citations from these and other computer-generated articles by other fictional scholars, Ike Antkare rose to the twenty-first position of most cited scientists on Google’s Scholarometer, beating Albert Einstein in the process.

Burkhard Morgenstern describes possibly the most hilarious prank involving the creation of fake scientists (with fake bios) to join editorial boards of spammish journals. He first became “Peter Uhnemann” (a character created by a German satirical magazine to troll a conservative politician), whose bio listed his research interest as “oximological microbiology, nonlinear submorphological endosaccharomorphosis, and applied endoplutomomics.” After Dr. Uhnemann was welcomed on a variety of editorial boards, Morgenstern decided to raise the bar by mobilizing “Hoss Cartwright” from the television show *Bonanza*. Dr. Cartwright was immediately welcomed on the editorial board of the *Journal of Primatology*, no doubt because of an impressive vitae featuring his affiliation with the Ponderosa Institute of Bovine Research, which he joined after extensive training in “Dunnowhat,” a postdoctoral fellowship at “Cowboy University” followed by a second postdoc at “Some shitty place in the middle of nowhere.” His last listed position was “senior cattle manager.”

In a different, more serious register, independent organizations—like CrossRef and Sideview—also work to critique gaming metrics in scholarly publishing. Not keen to make metrics history, Jennifer Lin describes the successes, future potential, and challenges of “altmetrics”—a large family of indicators that map references to publications and statements well beyond journal downloads, citations, and so on. It also does so across different media and platforms (including social media) rather than just journals or books. Perhaps because it is not one thing but many, and because it is rarely used by universities to evaluate faculty performance, altmetrics has not shared the criticism leveled against traditional metrics, nor has it provided widespread incentives for its gaming.[2] Furthermore, because of its complexity, the gaming of altmetrics “requires manipulating measurements across a diverse set of independent platforms [which] involves extensive coordination of multiple methods specific to each metric.” So far so good, but what will happen when altmetrics develops a more clearly defined market and a specific set of users and uses, like today's impact factor?

Elizabeth Wager too targets the downsides of metrics, but sees journals as part of the problem, not the solution. She proposes—simply and radically—to terminate the use of journal articles as the canonical genre of research dissemination—a convention that, while seemingly set in stone, is in fact nonobvious. The standard length of an article is unlikely to match the nature and scope of the topic, thus creating both "salami-sliced" and redundant publications. Also, the expectations of the genre tend to disincentivize the publication of large datasets and trials. Finally, the growing pressure to publish in high-impact journals may lead to hyping the importance of one's claim, or even to misleading reporting and data manipulation. Focusing specifically on data-intensive research, Wager argues that all these problems could be avoided by switching to research reports (rather than articles) structured in a way to make their data machine readable. Operating outside of any regime of metrics, these would be "outputs" rather than "articles"—outputs to be used and built upon rather than evaluated (in the sense that metrics give to evaluation). Of course, articles discussing and commenting on published data would still be produced, but they would not be the "main course" of science publishing anymore, thus forcing universities and government agencies to articulate "new methods for measuring research productivity."

Notes

1. Ivan Oransky, "Lawsuit Involving PubPeer Unmasks Commenter as Pseudonymous Whistleblower Clare Francis," *Retraction Watch*, April 13, 2015. http://retractionwatch.com/2015/04/13/lawsuit-involving-pubpeer-unmasks-commenter-as-pseudonymous-whistleblower-clare-francis/ (accessed October 10, 2016).

2. But there are exceptions, like Yves Gingras, *Bibliometrics and Research Evaluation: Uses and Misuses* (Cambridge, MA: MIT Press, 2016), pp. 67–68.

10

Retraction Watch: What We’ve Learned and How Metrics Play a Role

Ivan Oransky

Adam Marcus and I founded Retraction Watch (http://retractionwatch.com) in August 2010 for two reasons: As longtime journalists, we often found that retraction notices were opaque. And sometimes opacity was the best you could hope for; often, notices were misleading or even wrong. We also found that there were great stories behind retractions.

We have our own metrics at Retraction Watch, mostly just having to do with traffic to the site each month; we now have, on average, 150,000 unique visitors, and half a million page views. (However, we are not beholden to these metrics, as our revenue does not depend on advertising; we have at various times had generous funding from three foundations, and other income streams including freelance writing fees.) In terms of more traditional metrics, I can say we have been cited in the literature more than a hundred times. That means that if a blog could have an H index, we would have a good one. And it does not hurt when we talk to funders about the impact we are having on publishing practices and transparency.

Retraction Watch posts often begin with a tip—mostly a notice of retraction. But we also receive long emails from frustrated researchers, who have been laboring to correct a perceived wrong for months, if not years. We empathize and sympathize with their frustration—it is incredibly hard to get papers retracted from the literature, or even corrected or noted in some way.

As an illustration, take a piece by nutrition researcher David Allison and colleagues that appeared in *Nature* (Allison et al., 2016). They scanned the nutrition literature and found more than two dozen papers that they thought were deeply problematic. And they kept a pretty high bar. You can judge for yourselves, but if you look at the kinds of problems they were looking at in these papers, it was pretty clear something needed to be done. In a few cases, the journals retracted the paper, or published a letter from Allison and his team critiquing the findings, but in many cases the journals did nothing. It was very, very frustrating. We wrote a commentary on this eye-opening article (Oransky and Marcus, 2016) and a Q&A with the authors as well (McCook, 2016a). And although retraction is not always the best way to correct the scientific literature—corrections and correspondence, for example, may serve the purpose—the nuclear option is sometimes necessary. But it is very, very difficult to get a paper retracted. That is for a number of reasons, but one is clearly the stigma attached to retractions—although I will present some surprising findings on that in the conclusion of this chapter—and the fact that publishing papers in certain journals is the only way to earn grants, tenure, and promotions. They are, in a word, the metrics. Researchers will do anything to publish papers in some journals, including even creating fake authors (Marcus and Oransky, 2016). The opposite side of that coin is that many authors are very reluctant to admit to flaws in their work, and, by extension, to retract. I would put Allison and his colleagues in a "still relatively patient and professional" category in terms of their approach to trying to correct the record. But there are others who let their exasperation show. Some of our frequent commenters fall into that category, and they often find it difficult to maintain their equanimity. It can be very frustrating to try and correct the literature. Often well-meaning researchers who care about the public record use their free time to correct it, but it does not go well and they are vilified. To put it another way, whistle-blowers often fare far worse than those they are accusing of misconduct—and most problematic papers do not even earn an Expression of Concern, let alone a retraction. In short, the most common outcome for those who commit fraud is: a long career.

There is a really robust source of these conversations about the literature: PubPeer (Barbour and Stell, this volume, chapter 11). The site became a great resource for us, because we had a place to refer all the readers who were sending us allegations that we are not equipped—in terms of time, resources, and expertise—to properly vet and write about. PubPeer comments are a great resource for us, but so are the comments on our own site, in response to particular posts. There is much debate over comments and whether or not commenters should be allowed to remain anonymous (Blatt, 2015). We have chosen to allow commenters to post anonymously, because of the hierarchy of science and the dangers of speaking out against the powerful. But we moderate each comment heavily, and won't post any allegations that cannot be supported by evidence.

Here is an example of why we believe in the importance of anonymous comments. In 2014, we learned of a series of retractions from diabetes researcher Cory Toth, who used to be at the University of Calgary. We reported on them, and he actually said some very interesting things. “I am significantly apologetic, remorseful, and embarrassed that this occurred under my watch,” he told us. “Please know that I will not be publishing in the world of science in the future” (Marcus, 2014). The university had done an investigation, and they found misconduct, so he left.

Once we posted about four retractions, things became even more interesting. Readers started leaving comments that said, basically, “Oh, what about this paper that wasn’t retracted? You should look at this other figure. It looks a little bit odd. Something else here, there, and everywhere.” The university—unbeknownst to us—saw these comments and followed up on them. They reopened their investigation, and he ended up with five more retractions. And when Margaret Munro, then a reporter for Postmedia, called the university, they told her it was because of the Retraction Watch comments (Munro, 2014).

That story, and PubPeer, are important reminders that we are part of an ecosystem. What happens on PubPeer can result in a retraction (Keith, 2015). I can point to dozens of cases we have covered where the allegations first appeared in PubPeer. But it is not just us and PubPeer. There are many people out there focusing on similar issues. Some, like Science-Fraud.org, have been forced to shutter because of legal threats (Brookes, this volume, chapter 13; Oransky, 2013). And PubPeer faced legal action brought by a scientist who lost a job offer at least in part due to comments on the site. The scientist wanted PubPeer to unmask a commenter, but with the help of the American Civil Liberties Union, earned a judge’s ruling that allowed them to refuse (McCook, 2016b).

While we face those kinds of threats, too, we have yet to be sued, which we attribute to having worked in journalism for more than a decade and having developed a clear sense of libel laws and how they apply to our efforts. Also, because we are journalists, rather than practicing scientists, we can write critically of researchers, publishers, and institutions without fearing reprisals from those in power that would damage our own careers. There are others in our ecosystem, like Ben Goldacre, who are going from strength to strength. Many of us have been funded by the Arnold Foundation, whose support of this area should be acknowledged. It is very important if you get involved in this work to think of yourselves as part of an ecosystem.

To put it another way, there are lots of people out there who are as obsessive about their niches as Adam and I are about retractions. We are eager to host those fellow obsessives, in guest posts—such as a post about the Collaboration Score from its creator, Sarah Greene (Greene, 2016).

What did we learn in our first five years? One clear trend is that the retraction rate is on the rise. The number of retractions has also gone up, not surprisingly, as there are more papers. From 2001 to 2010, as very ably demonstrated in *Nature* by Richard Van Noorden, the number of retractions increased from about forty a year to four hundred. The number of papers only went up forty-four percent (Van Noorden, 2011). The number of retractions went up again to somewhere between five hundred and six hundred, and in fiscal year 2015, it was close to seven hundred, according to PubMed. And there are more retractions that are not captured by typical databases. (Keep in mind that despite the increase, the rate of retraction remains quite small—less than 0.05 percent of all papers.) Our database of retractions, made available in October 2018 at retractiondatabase.org years after the conference from which this chapter originated, demonstrated that those trends continued but that the rate of retraction may be leveling off (Brainard and You, 2018).

Does this mean that fraud is on the increase? We do not know, but we are certainly better at catching it. The introduction of plagiarism detection software changed a lot, as it enabled anyone to scan the millions of online papers. It is like the introduction of a new screening test: you would expect to see an increase in diagnoses once you start screening. At the same time, we have some evidence that misconduct of at least one kind—inappropriate image manipulation—is on the rise, thanks to a painstaking effort by Elisabeth Bik to scrutinize the Western blots in more than twenty thousand papers. Bik and her colleagues found that one in twenty-five papers contained problematic images—and that the rate of such issues had grown dramatically since 2000 (Bik, Casadevall, and Fang, 2016).

These are some of the reasons papers are retracted, in a fairly random order:

- **Plagiarism**: This is responsible for about ten percent of retractions.
- **Duplication**: You cannot really plagiarize yourself, but you can duplicate your own work. That is probably about fifteen percent of retractions.
- **Image manipulation**: Photoshopped photos of Western blots are a common reason for retraction, and a frequent source of comments on PubPeer.
- **Faked data**: Diederik Stapel may be the most well-known recent case of this, but it is not uncommon among reasons for retraction.
- **“For legal reasons”**: This one is sort of strange to us because it suggests that all the notices that do not include that term are illegal.

- **Fake peer reviews**: This is a trend we have been reporting on since 2012. It is relevant to a discussion of metrics because you can manage to have your paper accepted if you do your own peer review, taking advantage of the way editorial management systems are set up (Ferguson, Marcus, and Oransky, 2014). More than six hundred papers have now been retracted for that reason. And the fact that researchers do this in response to "publish or perish" pressures suggests a direct link to metrics.
- **Honest error**: This is responsible for about twenty percent of retractions, and when we see authors going out of their way to correct the record, we call attention to it by adding the case to our "doing the right thing" category.
- **Publisher error**: This is not a particularly important reason for retraction, and not all that common, but sometimes publishers print the same paper two or even three times. Apparently, the only way they know to correct the record is to issue a retraction, which means an author has a retraction—and the stigma that comes with it—for no fault of his or her own.
- **Authorship issues**: These can be a real headache for journals and universities. We have seen cases in which authors appear on papers they had nothing to do with, and other cases in which rightful authors are not named.
- **Lack of reproducibility**: This is controversial. Most scientists we speak to argue against retracting a paper just because later work overturns it, but some decide to retract their own papers for this reason.

Overall, two-thirds of retractions are due to misconduct, as Fang, Steen, and Casadevall showed in 2012. Sometimes, however, it is hard to tell. Here is a notice that was typical in a particular journal until quite recently: "This article has been withdrawn by the authors." That is not very helpful, but the *Journal of Biological Chemistry*, where the "article has been withdrawn by the authors" retractions appeared, must have become tired of having us beat them up after five years, so they changed their policy and now include details (Guengerich, 2015).

Speaking of metrics, here is how they might be related to how long it takes to retract. The average is about three years (Steen, Casadevall, and Fang, 2013). To be a conspiracy theorist for a moment, it is worth noting that three years is a year longer than the amount of time citations count toward a journal's impact factor. That means if journals can drag out the process, they would not take an impact factor hit. Similarly, authors and universities can drag out the process to make it less likely that a retraction will affect promotion or funding decisions.

And retracted papers keep being cited, often as if they had never been retracted (Budd, Coble, and Anderson 2011). As an illustration, we have a leaderboard of the top ten cited retracted papers (Retraction Watch, 2016b). Number two on the list at the time of this writing was Andrew Wakefield's infamous *Lancet* paper claiming a link between autism and vaccines (Wakefield et al., 1998). (Look, I just cited it.) And number one has had far more citations after it was retracted than before. Whether those are positive or negative citations, we do not know.

Which journals retract most? It turns out that journals with the highest impact factor—there are those metrics again–also have the highest rate of retraction. Again, we think that is mostly due to the fact that there are more eyeballs on those journals, although it is certainly possible that scientists are pushing the envelope to publish in those journals in ways that would constitute misconduct. "Meta-scientists" continue to debate the data on whether "publish or perish" plays a significant role in misconduct (Fanelli et al., 2017). And which researchers retract most? We like our leaderboards and our rankings at Retraction Watch. The top retraction holder has 183 (Retraction Watch, 2016a).

Let me leave you with an interesting finding that relies on metrics. A study—whose findings have been replicated by another group (McCook, 2015)—found that when people retract papers for fraud, you see what you would expect: their citations drop (Lu et al., 2013). In fact, citations in their whole subspecialty drop by about ten percent to fifteen percent. That is bad news, which, again, you would expect. But if you retract for honest error, however, and it is clear in the notice that the mistake is the result of honest error, you do not see that drop. So there is an example of using citations—a metric—to figure out what happens after a retraction. And it is also good news, in that good behavior is not punished.

Metrics: we just cannot get away from them.

References

Allison, David B., Andrew W. Brown, Brandon J. George, and Kathryn A. Kaiser. 2016. "Reproducibility: A Tragedy of Errors." *Nature* 530:27–29.

Bik, Elisabeth M., Arturo Casadevall, and Ferric C. Fang. 2016. "The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications." *mBio* 7(3):e00809–e0080916. doi:10.1128/mBio.00809–16.

Blatt, Michael R. 2015. "Vigilante Science." *Plant Physiology* 169(2):907–909.

Brainard, J., and J. You. 2018. “Rethinking retractions.” *Science* 362(6413): 390-393.

Budd, John C., Zach C. Coble, and Katherine M. Anderson. 2011. “Retracted Publications in Biomedicine: Cause for Concern.” In *Association of College and Research Libraries Conference*, pp. 390–395.

Fanelli, Daniele, Rodrigo Costas, Ferric C. Fang, Arturo Casadevall, and Elisabeth M. Bik. 2017. “Why Do Scientists Fabricate and Falsify Data? A Matched-Control Analysis of Papers Containing Problematic Image Duplications.” *bioRxiv*. http://biorxiv.org/content/early/2017/04/12/126805.

Fang, Ferric C., R. Grant Steen, and Arturo Casadevall. “Misconduct Accounts for the Majority of Retracted Scientific Publications.” *Proceedings of the National Academy of Sciences* 109(42):17028–17033.

Ferguson, Cat, Adam Marcus, and Ivan Oransky. 2014. “The Peer-Review Scam.” *Nature* 515(7528):480–482. doi:10.1038/515480a.

Greene, Sarah. 2016. “A New Metric: The Rapid Science Collaboration Score.” *Retraction Watch*. February 5. http://retractionwatch.com/2016/02/05/a-new-metric-the-rapid-science-collaboration-score/.

Guengerich, F. Peter. 2015. “Some Changes in Submission and Handling Policies at the JBC.” *Journal of Biological Chemistry* 290(46):28018–28019. doi:10.1074/jbc.E115.000001.

Keith, Ross. 2015. “Mirror Image in Plant Study Flagged on PubPeer Grows into Retraction.” *Retraction Watch*. September 25. http://retractionwatch.com/2015/09/25/mirror-image-in-plant-study-flagged-on-pubpeer-grows-into-retraction/.

Lu, Susan Feng, Ginger Zhe Jin, Brian Uzzi, and Benjamin Jones. 2013. “The Retraction Penalty: Evidence from the Web of Science.” *Scientific Reports* 3:3146. doi:10.1038/srep03146.

Marcus, Adam. 2014 “Pain Study Retracted for Bogus Data Is Second Withdrawal for University of Calgary Group.” *Retraction Watch*. April 7. http://retractionwatch.com/2014/04/07/pain-study-retracted-for-bogus-data-is-second-withdrawal-for-university-of-calgary-group/.

Marcus, Adam, and Ivan Oransky. 2016. “Why Fake Data When You Can Fake a Scientist?” *Nautilus*. November 24. http://nautil.us/issue/42/fakes/why-fake-data-when-you-can-fake-a-scientist.

McCook, Alison. 2015. “The Consequences of Retraction: Do Scientists Forgive and Forget?” *Retraction Watch*. June 2016. http://retractionwatch.com/2015/06/16/life-after-retraction-in-many-cases-its-forgive-and-forget/.

McCook, Alison. 2016a. “Want to Correct the Scientific Literature? Good luck.” *Retraction Watch*. February 3. http://retractionwatch.com/2016/02/03/six-barriers-to-correcting-the-scientific-record-qa-with-david-allison-and-andrew-brown/.

McCook, Alison. 2016b. “PubPeer Wins Appeal of Court Ruling to Unmask Commenters.” *Retraction Watch*. December 7. http://retractionwatch.com/2016/12/07/pubpeer-wins-appeal-court-ruling-unmask-commenters/.

Munro, Margaret. 2014. “Prolific University of Calgary Doctor Resigned After Research Team Caught ‘Manipulating’ and Faking Data.” *National Post*. September 8. http://news.nationalpost.com/news/canada/prolific-university-of-calgary-doctor-resigned-after-research-team-caught-manipulating-and-faking-data.

Oransky, Ivan. 2013. “Facing Legal Threats, Science Fraud Temporarily Suspends Posting.” *Retraction Watch*. January 2. http://retractionwatch.com/2013/01/02/facing-legal-threats-science-fraud-temporarily-suspends-posting/.

Oransky, Ivan, and Adam Marcus. 2016. “Why Is It So Damn Hard to Get a Paper Retracted?” *STAT.* February 3. https://www.statnews.com/2016/02/03/paper-retracted/.

Retraction Watch. 2016a. “The Retraction Watch Leaderboard.” Accessed May 15, 2016. http://retractionwatch.com/the-retraction-watch-leaderboard/.

Retraction Watch. 2016b. “Top 10 Most Highly Cited Retracted Papers.” Accessed May 15, 2016. http://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/.

Steen, R. Grant, Arturo Casadevall, and Ferric C. Fang. 2013. “Why Has the Number of Scientific Retractions Increased?” *PLoS One* 8(7):e68397. doi:10.1371/journal.pone.0068397.

Van Noorden, Richard. 2011. “Science Publishing: The Trouble with Retractions.” *Nature* 478:26–28. doi:10.1038/478026a.

Wakefield, Andrew J., Simon H. Murch, Andrew Anthony, John Linnell, D. M. Casson, Mohsin Malik, Mark Berelowitz, Amar P Dhillon, Michael A Thomson, Peter Harvey, Alan Valentine, Susan E Davies, and John A Walker-Smith. 1998. “Ileal-Lymphoid-Nodular Hyperplasia, Non-Specific Colitis, and Pervasive Developmental Disorder in Children.” *The Lancet* 351(9103):637–641.

11

PubPeer: Scientific Assessment Without Metrics

Boris Barbour and Brandon M. Stell

"She has a *Nature* paper!" How many times have you heard that sentence? And how many times was it accompanied by a discussion of what was in the paper? Maybe not so often, which is symptomatic of the problems in scientific research today. Worse, the obsession with the journal, at the expense of the result, is a pragmatic adaptation to the incentive structure of research today. Success in the competitions for jobs and grants will follow that publication, independently of its substance (or lack of it).

Publishing in *Nature* or one of several "glamour" journals is supposed to be synonymous with high-quality research of great significance. It represents a metric—a shorthand representation of the quality of research output that is often given numerical form as the "impact factor," described elsewhere in this volume. The lure of metrics is incredibly strong, because they offer an apparently efficient and objective basis for allocating resources such as grants and jobs, for which competition is fierce.

But every metric distracts our severely limited attention from the substance of the research. And, in doing so, metrics allow career success (or survival) to become dissociated from research quality. Defenders of the most sought-after journals will stress that stiff competition, stringent criteria for general interest, and attentive refereeing ensure that only the highest quality papers are accepted. Yet there are numerous examples of papers of dubious quality and with overblown interpretations being accepted in even the best journals. In many ways this is hardly surprising, because the true scientific impact of a paper is extremely difficult to judge when it is new.

The "top" journals distort science in other, more subtle, but equally pernicious ways. Their criteria for acceptance require papers to present "revolutionary concepts." Of course revolutions sometimes happen, but the reality is nearly always more prosaic. This means that aspiring authors feel pressured to adapt their presentations to fit these criteria, quite often exaggerating interpretations and misleading readers. Even worse, our whole approach to research has become governed by the desire to strike gold, even if most risky bets simply don't pay off and authors are obliged to spin their often useless results in some misleading way. All manner of solid, incremental, and useful work is deemed of too low impact to stand a chance of reaching a top journal and is therefore neglected by the grant evaluation committees and hiring committees in charge of identifying and nurturing the most promising research. We thus find ourselves in a situation where all of our research is constrained by the need to create whatever the high-impact journals consider to be high-impact stories. We are caught in a self-perpetuating game where the score is a publication that suits one of the top journals.

It is often stated that "science is self-correcting." If it is, then maybe the occasional excess won't be too damaging? However, as we have argued in our blog post "A crisis of trust,"[1] we believe that a combination of the pressures to publish high-impact research, the acute competition in research careers, and multiple levels of conflicts of interest have created an environment today that is hostile to any form of self-correction.

The traditional form of correction is the publication of a new paper, presenting contradictory findings and explaining how the original study was in error. But such studies are vanishingly rare today. Probably the major discouragement is editorial policy, because although the top journals often seem happy to accept somewhat shaky work raising an interesting possibility, they are almost never keen to publish a paper ruling out that sexy conclusion they previously published. There is a conflict of interest here, because the journal editors obviously prefer to avoid even implicit criticism of their evaluation process and brand.

A further obstacle is financial. Obtaining funding is extremely competitive and most funding bodies seek to identify the "revolutionary" work that will be publishable in the top journals. Proposing to redo somebody's experiments will in most cases be laughed out of the committee. A rare but hopeful counter-example was the large-scale project to replicate one hundred psychology experiments under the auspices of the Open Science Foundation, funded by the Laura and John Arnold Foundation.[2] Famously, fewer than one-half of the experiments replicated. Replications often require human resources as well as financial resources, but if it is neither possible to hire somebody to do the work nor to offer them a realistic chance of publishing in a top journal, it becomes doubly difficult to motivate people to carry out the replication. A final obstacle is social. Replicating work is almost uniformly seen as a sign of mistrust (it often is), and this can rapidly lead to poor relations with the original researchers and be perceived as aggressive by the community. Replicators risk exposing themselves to reprisals via the many anonymous decisions governing their careers (reviews of papers, grants, job applications, promotions).

An alternative and more lightweight method of correcting science is through correspondence and commenting. Indeed, issues in papers often don't require additional experiments or complete new manuscripts to be identified; such issues could include errors of experimental design, analysis, or interpretation. However, current journal systems for correspondence and for commenting have proved almost totally ineffectual. In the case of correspondence, many journals simply do not have a correspondence section. For those that do, it remains extremely constrained in terms of the number of items a journal will publish and the allowed size of those items. Less apparent, but equally inhibiting, editorial policies repress or dissuade most if not all attempts to publish correspondence. At times, the discouragement and obstacles faced by correspondents are almost comical. Although there should be no space limitation for commenting on journal websites (at least for those that offer the facility), many of the same discouragements operate. In sum, there is a widespread and self-fulfilling perception that journals do not welcome correspondence or comments that criticize their publications. The result is that correspondence and commenting on journal websites make no significant contribution to the correction of science.

Some journals have shown recent signs that they are becoming aware of this problem, but, as set out in our blog "Nature editors: all hat and no cattle,"[3] even thought leaders such as *Nature* are struggling to develop a coherent "correction" strategy within the constraints of their business model and presumptions of excellence. At the time of writing, the journal *Circulation Research* has formulated a promising policy, in which they state that refutations will be considered to have the same impact as the original work they address. This would remove one huge barrier at glamour journals, which almost invariably have judged refutations to fall below their general interest thresholds.

Thus, the present combination of career incentives and journal editorial policies has led to a situation where most researchers actively avoid criticizing directly or indirectly any published work. The self-correction of science is sick, although there are now some signs of progress.

It was in this context that PubPeer was developed and launched almost five years ago. The goal of PubPeer is to offer a web platform (https://pubpeer.com) for centralized discussion of publications, in order to create a community of "post-publication peer reviewers" and offer a new mechanism of scientific assessment. The website is independent of the journals, removing many of the psychological barriers and conflicts of interests inhibiting discussion and criticism of papers. Shortly after our launch, we introduced a key feature—-strong user anonymity—-and from that point on our comment volume climbed rapidly. It has become clear that a large number of substantive and critical comments had never before found an outlet. Specifically, a very large number of comments (concerning thousands of papers) have been posted that highlight manifest misconduct in the preparation of figure data. Typical issues include fabrication of western blot data through duplication of bands. Through observing these numerous cases, we have come to believe that the suppression of self-corrective mechanisms in science has enabled a surprisingly large number of researchers to build very successful careers based upon the most dubious of research practices.

Anonymity is PubPeer's most controversial aspect, but we remain convinced that its benefits strongly outweigh its disadvantages. (Our analysis of this is laid out in our blog "Vigilant Scientists.")[4] Although anonymity apparently introduces the possibility for people to be defamed with impunity, there are in fact layered safeguards to minimize this. First and foremost, our guidelines make clear that comments must be based upon public facts and/or logic, such that other readers can verify each comment's conclusions for themselves; hearsay is forbidden. These guidelines are enforced by moderation and reporting facilities for all posted comments. Of course, there are additional guidelines regarding politeness in posts, avoiding ad hominem arguments and discouraging speculation. These guidelines, combined with the constraints of the scientific discourse, which is inherently fact based, have, we believe, enabled a very high accuracy of user comments and allowed the site to avoid most of the problems associated with anonymous commenting elsewhere on the internet. In addition, authors are alerted to comments and encouraged to respond to any questions or criticisms by defending and explaining their work. Many of the issues raised on PubPeer would be solved instantly by posting the original data, yet most authors have not chosen this course of action.

Most of the arguments against PubPeer, and more specifically its anonymity, have focused on the belief that it will enable unfounded denigration of researchers. However, the site's critics are rarely able to produce even a single example of a career that has been unjustly harmed by criticism on PubPeer, despite the large number of comments now in the PubPeer database. For sure there are researchers who commented on PubPeer whose careers have been harmed, but in all of the high-profile cases that we are aware of, the criticisms have been found to be accurate and justified. In truth, only criticisms that are valid are damaging.

Against this perhaps surprisingly slight "defamatory" risk, we balance the benefit of the strong encouragement of post-publication peer review by anonymity on PubPeer. The kind of information posted to PubPeer can benefit readers in a number of ways. Comments typically highlight research that is unreliable in some way. Knowledge of this can prevent researchers wasting their time and (taxpayers') money attempting to build upon some apparently exciting but flawed breakthrough. Furthermore, since public policy and medical guidelines are, at least in theory, based upon evidence, often in the form of published research, early warning of potential problems can guard against erroneous policy and even save lives. The guiding policy at PubPeer can be summarized as putting first the readers and users of published research, even if that can sometimes be uncomfortable for the authors of the publications discussed. As we stated on our blog: "…a few ruffled academic feathers pale into insignificance when patients' lives, taxpayer billions, and young researchers' careers are at stake. We also suspect that the researchers' employers—those same taxpayers and patients—would share this point of view." We also believe that defenders of authors' rights not to be criticized have forgotten that publication is a choice, freely made by the authors. To put things bluntly, authors who don't wish their work to be criticized or questioned are always free not to publish.

At the time of writing, PubPeer is approaching five years of operation. It is clearly here to stay and is exercising a growing influence on research, both through a number of spectacular scandals that have been brought to light via the site, but also via a cumulative effect on journals, institutions, and researchers themselves, as the site becomes integrated into research, editorial, and administrative workflows. Although no journals or databases currently integrate PubPeer comments, users can install browser plugins that provide as-you-browse alerts to papers with comments.

With those five years of experience, what has emerged is that most comments are indeed critical. Clearly the most effective motivation to comment is when a reader perceives a problem in a paper. In a way this is unsurprising: firstly, science proceeds by falsification; secondly, most papers are of necessity written to present the work in the most positive light possible. Decline and decay is the fate for many papers.

Post-publication peer review, which PubPeer facilitates, is compatible and synergistic with two other key trends refocusing our attention on the substance of research. The first is the use of preprints; servers such as the ArXiv and bioRxiv offer instant access to papers as soon as they are ready, eliminating a delay of many months and even years. Arguments about low-quality papers no longer being filtered by peer review can be countered by the extremely successful example of the ArXiv and by the flood of journals without meaningful quality controls. We recall moreover that even the top journals are clearly susceptible to allowing "impact" to compensate for rigor in the papers they publish. PubPeer accepts comments on preprints for both ArXiv and bioRxiv and would implement others if and when they come online (we are however personally in favor of a few, centralized preprint servers).[5]

The second trend is toward full data access. Several publishing groups (including *PLoS* and *NPG* at the time of writing) require authors to share their data upon request or even before publication the full set of their original data. Although the systems and processes of this are by no means reliable or routine, the direction is clear. Access to the data will enable much more meaningful post-publication review and discussion. Authors and journals will no longer be able to suppress checks and reanalyses by controlling (usually refusing) access to the data. One consequence of this will be that many of the kinds of issues that appear on PubPeer and currently lead to somewhat ungracious and inconclusive discussions will in future be resolved directly, by access to the original data that is today withheld. In some respects, we are in a transition period.

What should the future of scientific research look like? What we would like to see and what we are working toward is an environment that is much more accepting of discussion, criticism, reanalysis, and replication. This in turn will focus attention on the substance of publications (whatever their format), with the consequence that authors will be under pressure to publish good quality work. This contrasts with the current situation where authors can largely escape criticism once a "definitive" publication has appeared. Intensive post-publication review also places pressure on authors to provide "after-sales service." Those that publish low-quality work will have to spend more time explaining and defending their work, which is as it should be.

Will we need journals in the future? Preprint servers could certainly satisfy the requirements of publication at a tiny fraction of the cost. We are quite unconvinced by journal claims of a high added value for their reviewing, formatting, and indexing.[4] Many publishers also try to argue that preprints would produce a flood of low-quality work and that researchers do not have time to trawl through millions of preprints to find those that are of high quality and of interest to them. Yet it is unlikely that the current system prevents publication of low-quality work (a bad paper eventually gets published in one of the more than ten thousand currently existing journals). Furthermore, physicists all use the ArXiv, which divides papers into broad subjects (an approach replicated by bioRxiv), and don't seem to have any greater difficulty in keeping abreast of new work than researchers in other fields. Once the community grows to a critical mass, post-publication review on sites such as PubPeer could replace (or complement) some if not all of the evaluation that is currently carried out with pre-publication review, although the volume of the latter obviously greatly exceeds post-publication review at this time.

Finally, we return to the issue of metrics, be they explicit (e.g., number of citations) or implicit (the journal). As we have argued above, we believe that all metrics are dangerous attempts to measure the unmeasurable, encourage gaming of the metric, and distract attention from the substance of the science. We consider it unlikely that any metric will prove beneficial to the management of science.

Notes

1. The PubPeer Foundation. 2015. "A crisis of trust." Accessed July 6, 2016. https://blog.pubpeer.com/publications/EE46E14F6AA97049928835DBC6B908.
2. Open Science Collaboration. 2015. "Estimating the reproducibility of psychological science." *Science* August 28, 2015: Vol. 349, Issue 6251, doi:10.1126/science.aac4716.
3. The PubPeer Foundation. 2016. "Nature editors: All hat and no cattle." https://blog.pubpeer.com/publications/AE11BE44CF3C40A558F3B453BF53C7.
4. The PubPeer Foundation. 2016. "Vigilant Scientists." Accessed July 6, 2016. https://pubpeer.com/publications/B6CF3DB974A8ECC64B1A0303BBCD6F.
5. Cell Press. 2016. "Let's talk about preprint servers." Accessed July 6, 2016. http://crosstalk.cell.com/blog/lets-talk-about-preprint-servers.

12

The Voinnet Affair: Testing the Norms of Scientific Image Management

Catherine Guaspare and Emmanuel Didier

In the 1980s, it was already obvious that large companies were using metrics to track numerically the efficacy and efficiency of their employees. Management methods based on quantification gained widespread recognition during this decade (Camp, 1989; Hammer and Champy, 1993). Since the 2000s, these metrics seem to have migrated far beyond the borders of the corporate world, permeating the public bureaucracies of most liberal states, giving rise to what has come to be called the "New Public Management" (Osborne and Gaebler, 1992). Education was one site where these new metrics were first applied. While their success proves that they have been useful to some constituencies, their use has also been accompanied by drawbacks and complaints, which have often focused on the gaming practices that metrics fostered (Bruno and Didier, 2013). There are now suggestions that the problem may be worse than that. As universities and research institutions have adopted metrics of performance evaluation (Lawn, 2014), it has also become clear that the life sciences are plagued with widespread misconduct. Is there a correlation?

The literature offers some explanations for the apparent increase of misconduct cases (Steen, Casadevall, and Fang, 2013). One could argue that the barriers to publication have been lowered in the last twenty years or so, thus letting more scientists publish fraudulent articles. But this argument has limitations. First, it does not explain why the barriers to publications have been lowered. Second, over the same period, we have witnessed a series of public scandals triggered by findings of misconduct. Given that one of the social functions of scandal is precisely to *enforce* norms (Dampierre, 1954; Molotch and Lester, 1973; Boltanski, Claverie, and Offenstadt, 2007; Adut, 2008), we should not expect to find scandals if there was a simple lowering of the quality requirements in science. A different explanation is offered by the scientists themselves who often complain that they are under pressure, which may tempt them to produce subquality science. But the problem with the word "pressure" (which, by the way, is shared with many other professions from the police to medical doctors [Bruno and Didier, 2013]) is that it is mainly a complaint, a denunciation. It is not analytical.

Our hypothesis is that certain dimensions of the practice of science have experienced important transformations in the last twenty years, caused in part by the performative effects of metrics. These innovations have blurred the previous norms regulating scientific practices. At the same time, a new ecology (Star, 1995) of watchdogs and fraud detection has emerged that gives increased visibility to cases of misconduct, which in turn stir public controversies or "affairs" through which norms are re-instantiated.

Rethinking Misconduct Through High-Visibility Affairs

The practice of science has changed in the last twenty years. The digitalization of scientific images and the accessibility of image-processing software (such as Photoshop) have completely reshaped the way images are used and processed in the laboratory, posing questions about the acceptable limits of image manipulation (Rossner and Yamada, 2004). At the same time, changes to the way laboratories are managed (caused by the increased size of research teams and changes in the funding structure) have deeply transformed the role of Principal Investigators (PIs), who are now managers as much as researchers. In turn, this has changed the distribution of responsibility within the laboratory concerning the handling of the images (Frow, 2012).

Since science is a competitive community where all try to succeed, there are incentives to use new tools up to the limit, which sometimes leads scientists to lose the sense of where those limits are, especially in regard to new practices like digital image processing and use. The main scientific journals are stepping in trying to establish guidelines but, until now, they have not resolved all of the ambiguities that emerge from the use of these new tools (Frow, 2012). Practitioners, therefore, may not know for sure what are the limits that should not be crossed given that the community is still elaborating and stabilizing them.

Controversies and scandals are defining events in making social norms explicit (Boltanski, 1990; Bloor, 1991; Claverie, 1998). They are oppositional moments when ethical and epistemological standards become unstable and have to be newly decided upon, breaking up the scientific community into groups advocating for a certain set of norms against another set, forcing them to make their conception of the norms explicit. For example, some scientists have recently expressed a sentiment of injustice when witnessing practices that they found shocking but brought success to those who deployed them:

> The climate of distorted incentives has been exploited by some scientists to build very successful careers upon fabricated data, landing great jobs, publishing apparently high-impact research in top journals and obtaining extensive funding.… Honest scientists struggle to compete with cheats in terms of publications, employment, and funding. (PubPeer, 2016)

Some who felt this type of resentment did not simply express it but also helped establish a whole new ecology of watchdogs to scrutinize scientific production. And when they identify something disturbing or puzzling, they find ways to have the scientific community confront these issues. They are social actors that put science to a test.

To analyze the ways in which a specific scandal becomes the moment when norms are expressed, we focus on a recent well-known example. In 2014 and 2015, a set of publications authored or co-authored by a renowned French life scientist, Olivier Voinnet, became the target of a series of public critiques. They were posted on the website PubPeer, one of the main new watchdogs that allows scientists to anonymously discuss and criticize scientific papers (Boris Barbour and Brandon Stell, this volume, chapter 11).

Public Critiques and the Mediatization of the "Voinnet Affair"

In 2014, Voinnet was a forty-three-year-old, high-profile plant molecular biologist widely recognized for his work on RNA interference, a defense mechanism that allows plants, but also invertebrates and mammals, to fight viruses. After a thesis completed under the supervision of David Baulcombe (a major plant scientist and a pioneering researcher in gene silencing), Voinnet was recruited by the French National Center for Scientific Research (CNRS) as a permanent researcher and appointed at the "Institut de Biologie Moléculaire des Plantes" (IBMP) in Strasbourg. He was soon promoted to first-grade senior researcher, the highest position for a CNRS researcher. Since 2010, he has been on secondment at the Swiss Federal Institute of Technology in Zurich (ETHZ), where he has set up a new lab. Practically every year since 2004, he has won prizes (the Academy of Science Prize, the Liliane Bettencourt prize for Life Sciences, the EMBO Gold medal) that recognized his work and helped sustain his research and his lab. During these years, Voinnet benefited from competitive European and French grants. By 2014, he ran two labs (one in Strasbourg, and one in Zurich, with ten members in the former and thirty members in the latter) (ETHZ, 2015c) and had already published around one hundred papers, some in high-profile journals (including *Nature*, *Cell*, *Science*, *EMBO Journal*, and *PNAS*), and some of them highly cited.

In September 2014, images related to ten papers co-signed by Voinnet were flagged on PubPeer. Anonymous peers suspected that some images had been inappropriately modified. Some appeared too bright or too clean or had been manipulated without providing information about the process that had been used. On January 9, Baulcombe (who was the corresponding author on half of the suspected publications posted since September) left a statement on PubPeer for each of the flagged papers (which had grown to twelve in the meantime), explaining that he had been aware of the problems, had begun an investigation, and would notify the editor. One of those statements was signed jointly by Baulcombe and Voinnet. On the same day, *Retraction Watch* (Ivan Oransky, this volume, chapter 10), another watchdog, informed its readers of the ongoing allegations posted on PubPeer about Voinnet and Baulcombe's work (Oransky, 2015a). By so doing, the blog publicized these concerns to a much broader audience, opening new discussions that quickly spread with the comments following the post.

Voinnet, corresponding author of several of the papers, left a post on PubPeer echoing Baulcombe's earlier statement that he was aware of the problems, was investigating, and would notify the editors. As the discussion on both PubPeer and *Retraction Watch* picked up pace, it amplified the concerns, and the scrutiny. By the end of January, thirty-five papers co-authored by Voinnet from different stages of his career were discussed on PubPeer and 255 comments had been left on this website. *Retraction Watch* recorded about 120 comments under its first post about this case. One month later, on February 18, a first correction occurred for a 2014 paper in *Genes & Development*. The information was immediately relayed and publicized through comments on PubPeer and *Retraction Watch*.

The mediatization of this affair by newsmagazines or newspapers further expanded the audience of the case. At first, two articles by the science freelance journalist Leonid Schneider (who was also active on *Retraction Watch*) were released in January in the *Laborjournal* online (Schneider, 2015a) and, in March, in the *LabTimes* online (Schneider, 2015b). But an article published in *Le Monde* (France's most important newspaper) at the end of March (Morin and Larousserie, 2015a) brought the affair to a large national public. The article reported the ongoing suspicions concerning figure manipulation that were expressed on PubPeer and *Retraction Watch* about articles co-authored by Voinnet. Also, *Le Monde* highlighted for the first time the existence and role played by the two websites, PubPeer and *Retraction Watch*.

The affair took a new turn on April 1, with a comment left on PubPeer by Vicki Vance, a plant scientist at the University of South Carolina (PubPeer, 2015). She wrote that on three occasions she had been the peer reviewer of one of the papers produced by Voinnet's lab and subsequently flagged it on PubPeer. Based on her repeated experience, she became convinced that the authors had lied about several data presented in the article. This paper had been rejected by the first two journals that had asked her for a review. However, the third journal—*Plant Cell*—decided to publish it in 2004, despite Vance's statement in her report that some part of the text or some figures had been deliberately twisted or improperly manipulated to support the results (Vance, 2015). In comments on the websites, some stated that if the authors were at fault, then the journal shared a responsibility for publishing a dubious article that encouraged to continue these questionable practices. On April 9, *Le Monde* (Morin and Larousserie, 2015b) reported this new development by writing that Voinnet was now being "accused of lying."

The same day, the CNRS (CNRS, 2015a) and ETHZ (ETHZ, 2015a) officially announced that they had already opened an investigation after being aware of the potential image manipulation in a large amount of publications co-signed by Voinnet. Following their procedures, each institution had set up a commission made of experts to investigate the allegations posted on PubPeer that were then broadly disseminated by the media. Four days later, *Retraction Watch* in its third post about the case (Oransky, 2015b) relayed this development. It also reported that Vance had released her original peer review of the *Plant Cell* paper on ResearchGate (Vance, 2015) (which opened up a new controversy because peer review is expected to remain private), and that she had also sent an open letter to the CNRS and the ETHZ.

The first retraction occurred in the midst of these institutional investigations. On June 2, 2015, *Plant Cell* complied with a request for retraction by the authors (Plant Cell, 2015) concerning the paper that had been highlighted by Vance on websites and newspapers.[1] This retraction went against the claim of the CNRS and ETHZ in their press release that the results were not affected by the potential image manipulation. The

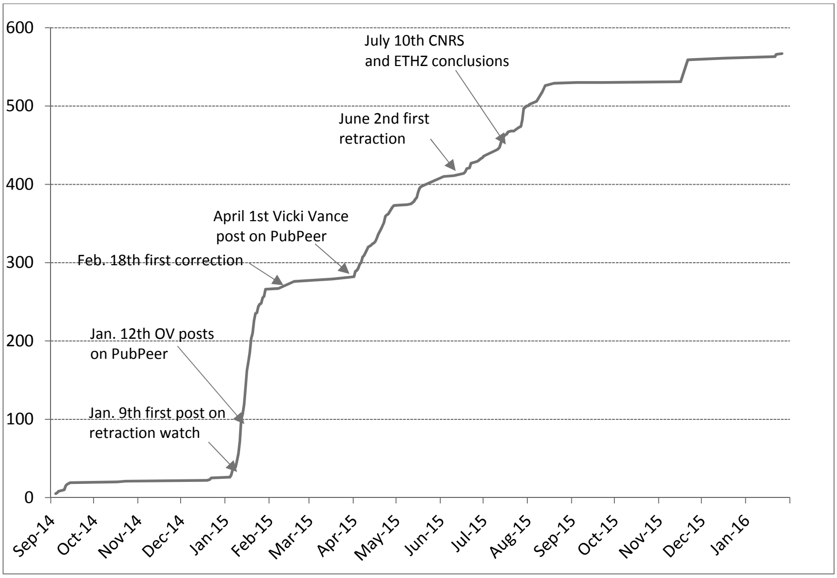


Figure 12.1
Cumulative frequency of posts on PubPeer about Voinnet's publications.

journalist from *Le Monde* underlined this in "A New Step in the Affair" (Morin, 2015). Meanwhile, the affair's visibility continued to grow, as shown by the number of posts on PubPeer.

The Institutions React

On July 10, two separate press releases (CNRS, 2015b; ETHZ, 2015b) published findings of the investigations carried out by the research institutions that employed Voinnet. The ETHZ reported that (ETHZ, 2015c)[2] while the commission did not find evidence of fraudulent fabrication of data in order to alter experimental results (a category 1 offence, the most severe in its own classification of research malpractice established for the case), various errors and manipulations with different levels of seriousness[3] were identified in the figures of about twenty papers. Images had been "beautified" by clearing the "background clouding," some bands were "duplicated," some figures were published without explanation about the processing they received, some papers shared the "same loading control images," and sometimes the same loading control images were used in different figures in the same paper. In other cases, the report pointed out that "mock idealized figures" produced for internal use during lab meetings had been "mistakenly used" in articles (ETHZ, 2015c). However, based on its assessment of evidence from lab notebooks and the raw images contained in them, the report stated that "the experiments reported in the investigated publications had been conducted and recorded carefully"—a statement that was then mirrored in the title of the ETHZ press release: "Conducted Properly—Published Incorrectly." Unlike the ETHZ, however, the CNRS considered that the errors and manipulations in Voinnet's papers constituted scientific misconduct (CNRS, 2015 b).

Both investigative commissions saw this series of data misrepresentations not as simple successions of random mistakes, but as the consequences of bad practices and bad habits in the presentation of scientific data. As head of the research group, Voinnet was supposed to supervise the treatment and presentation of the scientific results and also to guide his team, especially the junior scientists, to ensure the quality of their work. On this point, the commissions concluded that Voinnet did not meet his ethical and professional obligations. For both institutions, his responsibility was then not only as an author or co-author, but also as a team leader. Additionally, the ETHZ report implicitly pointed out the influence of metrics on these misconducts: "Being at the cutting edge of science, the OV laboratories were exciting, high-pressure, and fast-working, and being fast was part of the problem as OV himself admitted." The metrics might well be an element of this "pressure," leading up to these "speeding offenses," as this commission called them.

The two institutions imposed different sanctions on Voinnet. The president of the CNRS ordered him suspended from the CNRS for two years after his return from Zürich (CNRS, 2015c). The executive board of the ETHZ decided that Voinnet would "receive an admonition from the president of ETHZ" and that measures would be implemented to improve the working practices of the laboratory. Voinnet was also forced to relinquish his leading position at the CNRS laboratory in Strasbourg to focus exclusively on his ETHZ team. The asymmetry between these positions was partly related to the fact that, because Voinnet's employment by the CNRS and the ETH started on different dates, the two institutions did not have to investigate the same sets of papers. There were other sanctions. On January 2016, the Swiss National Science Foundation stopped Voinnet's funding (1.25 million Swiss francs) and banned him from its grants for three years. EMBO revoked his gold medal (Palus, 2016), and the affair led to the retraction of eight and the correction of twenty-four of Voinnet's papers since January 2015.[4]

Conclusion

To conclude, here are two points. First, the early 2010s saw the emergence of a new ecology of watchdogs, fueled by a sense of injustice that some scientists felt when seeing colleagues being rewarded for work they saw as flawed or even fraudulent. Through different means (such as anonymity to protect less established scientists, active online presence, and good connections with the press), they have been able to reach a broad audience when publicizing their serious doubts about the integrity of certain scientific works. They are now able to launch resounding "affairs," in public arenas so wide that government and academic institutions have to react and take a stance.

Second, there were image management practices apparently common (but kept discreet) in some laboratories that were instrumental to building successful careers. When exposed to the general public, however, these practices made those careers tremble. Driven in part by the wide public echo produced by the new watchdogs, institutions find themselves forced to confront these practices. As a result, they produce investigations, tools (such as new classification of misconducts), and organizations (such as the creation of a position of "scientific integrity adviser" by the CNRS)[5] to judge and slowly establish the material foundations of new norms for these practices.

Postscript

This article focuses on the inquiries for scientific misconduct led by the CNRS and the ETHZ in 2015. It is worth mentioning that in 2016 a new commission led by the CNRS assisted by the ETHZ was set up to investigate dubious figures into five publications from Voinnet's former CNRS lab at Strasbourg (ETHZ, 2016). In contrast with the previous findings, this last investigation revealed, alongside simple errors, evidence of intentional fabrication of data. The CNRS and the ETHZ cleared Voinnet of any unethical manipulation of data, but the CNRS holds him responsible for the repeated breach of good scientific practices that occurred under his supervision (CNRS, 2018; ETHZ, 2018).

Notes

1. In *Le Monde*, the journalists specified that Olivier Voinnet requested the retraction of the paper to *Cell* at the end of March (Morin and Larousserie, 2015b).

2. Contrary to the ETHZ, the CNRS did not make public the report of its commission of investigation.

3. Most of them fall into category 3 ("nonannotated processing of images") and some in category 2 (production of idealized figures to make them "more convincing without affecting the overall conclusion of the original experiment"). The report highlighted also that some papers might contain genuine errors (category 4, "unintended publication of erroneous images in place of the correct one").

4. According to the *Retraction Watch* database, consulted October 12, 2018. http://retractiondatabase.org/RetractionSearch.aspx#?auth%3dolivier%2bvoinnet.

5. http://www2.cnrs.fr/presse/communique/5737.htm.

References

Adut, Ari. 2008. *On Scandal: Moral Disturbances in Society, Politics, and Art*. Structural Analysis in the Social Sciences. Cambridge, New York: Cambridge University Press.

Bloor, David. 1991. *Knowledge and Social Imagery*. 2nd ed. Chicago: University of Chicago Press.

Boltanski, Luc. 1990. *L'Amour et la Justice Comme Compétences: Trois Essais de Sociologie de l'Action*. Collection Leçons de Choses. Paris: Métailié.

Boltanski, Luc, Elisabeth Claverie, and Nicolas Offenstadt, eds. 2007. *Affaires, Scandales et Grandes Causes: de Socrate à Pinochet*. 1 vol. Les Essais. Paris: Stock.

Bruno, Isabelle, and Emmanuel Didier. 2013. *Benchmarking: l'État Sous Pression Statistique*. Paris: Zones.

Camp, Robert C. 1989. *Benchmarking: The Search for Industry Best Practices That Lead to Superior Performance*. Milwaukee, White Plains, NY: Quality Press; Quality Resources.

Claverie, Elisabeth. 1998. "La Naissance d'une Forme Politique: L'affaire Du Chevalier de La Barre." In *Critique et Affaires de Blasphème à l'Époque des Lumières*, Philippe Roussin, 185–260. Paris: Honoré Champion.

CNRS. 2015a. "Allegations Concerning Olivier Voinnet's Publications: the CNRS Sets Up a Scientific Commission Of Inquiry." April 7. http://www2.cnrs.fr/en/2549.htm.

CNRS. 2015b. "Completion of the Procedure Against Olivier Voinnet." July 10. http://www2.cnrs.fr/en/2593.htm for the CNRS.

CNRS. 2015c. "Bulletin Officiel du CNRS." June–July, 386. https://www.dgdr.cnrs.fr/bo/2015/juin-juillet-2015.pdf.

CNRS. 2018. "Bulletin Officiel du CNRS." September, 382–383.

Cokol, Murat, Fatih Ozbay, and Raul Rodriguez-Esteban. 2008. "Retraction Rates Are on the Rise." *EMBO Reports* 9(1): 2. doi:10.1038/sj.embor.7401143.

Dampierre, Eric de. 1954. "Thèmes Pour l'Etude du Scandale." *Annale ESC*, IX, 3.

ETHZ. 2015a. "Allegations Under Investigation." April 7. https://www.ethz.ch/en/news-and-events/eth-news/news/2015/04/allegations-under-investigation.html.

ETHZ. 2015b. "Conducted Properly—Published Incorrectly." July 10. https://www.ethz.ch/en/news-and-events/eth-news/news/2015/07/investigation-result-voinnet.html for the ETH Zürich.

ETHZ. 2015c. "Report of the ETH Commission of Inquiry Set Up to Clarify Allegations Against Prof. Olivier Voinnet of ETH Zurich." June 17. https://www.ethz.ch/content/dam/ethz/news/medienmitteilungen/2015/PDF/untersuchungsbericht/Report_of_ETH_Commission_Voinnet.pdf.

ETHZ. 2016. "CNRS and ETH Are Setting Up a Commission of Inquiry." September 8. https://www.ethz.ch/en/news-and-events/eth-news/news/2016/09/commission-of-inquiry.html.

Frow, Emma K. 2012. "Drawing a Line: Setting Guidelines for Digital Image Processing in Scientific Journal Articles." *Social Studies of Science* 42(3):369–92.

Hammer, Michael, and James Champy. 1993. *Reengineering the Corporation: A Manifesto for Business Revolution*. New York: Harper Business.

Lawn, Martin. 2014. *Shaping of European Education: Interdisciplinary Approaches*. Studies in European Education. Florence: Taylor and Francis.

Molotch, Harvey, and Marilyn Lester. 1973. "Accidents, Scandals, and Routines: Resources for Insurgent Methodology." *Critical Sociology* 3(4):1–11. doi:10.1177/089692057300300401.

Morin, Hervé. 2015. "Un Article "Manipulé" d'une Star Française de la Biologie est Retiré." *Le Monde*. June 3. http://abonnes.lemonde.fr/sciences/article/2015/06/03/un-article-manipule-d-une-star-francaise-de-la-biologie-est-retire_4646446_1650684.html.

Morin, Hervé, and David Larousserie. 2015a. "Une Star de la Biologie dans la Tourmente." *Le Monde*. March 31.

Morin, Hervé, and David Larousserie. 2015b. "Olivier Voinnet, Star de la Biologie, Accusé de Mensonge." *Le Monde*. April 4.

Oransky, Ivan. 2015a. "Award-Winning Plant Researcher Correcting Several Papers Critiqued on PubPeer." *Retraction Watch*. January 9. http://retractionwatch.com/2015/01/09/award-winning-plant-researcher-correcting-several-papers-critiqued-pubpeer/.

Oransky, Ivan. 2015b. "Investigations into Voinnet's Work Announced; Critic Publishes Original Peer Review." *Retraction Watch*. April 13. http://retractionwatch.com/2015/04/13/investigations-into-voinnets-work-announced-critic-publishes-original-peer-review/.

Pain, Elisabeth. 2009. "Finding His Voice in Gene Silencing." *Science*. May 29.

Palus, Shannon. 2016. "EMBO Takes Back Voinnet's Award, Investigates Other Awardee Who Just Lost a Nature Genetics Paper." *Retraction Watch*. January 28. http://retractionwatch.com/2016/01/28/embo-takes-back-voinnets-award-investigates-other-awardee-who-just-lost-a-nature-genetics-paper/.

Plant Cell. 2015. Retraction: "Patrice Dunoyer, Charles-Henri Lecellier, Eneida Abreu Parizotto, Christophe Himber, and Olivier Voinnet.(2004). Probing theMicroRNA and Small Interfering RNA Pathways with Virus-Encoded Suppressors of RNA Silencing." *Plant Cell* 16:1235–1250. http://www.plantcell.org/content/early/2015/06/02/tpc.15.00407.full.pdf.

Politix. 2005. "A l'Epreuve du Scandale." n°71.

Osborne, David, and Ted Gaebler. 1992. *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector*. Reading, MA: Addison-Wesley.

PubPeer. 2015. "PubPeer > Plant Cell." January 5. https://pubpeer.com/publications/15084715.

PubPeer. 2016. "A Crisis of Trust | PubPeer." March 7. http://blog.pubpeer.com/?p=164.

Rossner, Mike, and Kenneth Yamada. 2004. "What's in a Picture? The Temptation of Image Manipulation." *The Journal of Cell Biology* 166(1):11–15.

Schneider, Leonid. 2015a. "Paper von Olivier Voinnet unter Beschuss." *Laborjournal* online. January 20. http://www.laborjournal.de/editorials/906.lasso.

Schneider, Leonid. 2015b. "Too Much to Be Nothing?" *Lab Times* online. March 24. http://www.labtimes.org/editorial/e_594.lasso.

Star, Susan Leigh (Ed.). 1995. *Ecologies of Knowledge. Work and Politics in Science and Technology*. Albany, NY: State University of New York Press.

Steen, R. Grant, Arturo Casadevall, and Ferric C. Fang. 2013. "Why Has the Number of Scientific Retractions Increased?" *PLoS One* 8(7):e68397. doi:10.1371/journal.pone.0068397.

Stevens, H. 2013. *Life Out of Sequence. A Data-Driven History of Bioinformatics*. Chicago: University of Chicago Press.

Vance, Vicki. 2015. "This Is My Review of the Voinnet 2004 Plant Cell Paper Downloaded from the Journal's Website." April 7. https://www.researchgate.net/publication/274637686_This_is_my_review_of_the_Voinnet_2004_Plant_Cell_paper_downloaded_from_the_journal's_website_The_review_is_from_November_2003.

13

Crossing the Line: Pseudonyms and Snark in Post-Publication Peer Review

Paul S. Brookes

To discuss the role of snarkiness in scientific criticism, I will start out with an example. It is fun to think of Johnny Cochran, the lawyer from the O. J. Simpson trial, and his famous quotation, *"If the glove doesn't fit, you must acquit."* I think we can adapt this for science, and say *"If the data is whack, you must retract."*

In addition to snarkiness, I will discuss pseudonyms and anonymity, and the fine line between what we can say to criticize each others' science, without getting into slander/libel and *ad hominem* attacks. This line is not well defined for scientists (for example, the difference between saying *"I think these figures are too sufficiently similar as to have occurred by pure coincidence"* versus *"these figures are identical"*). The blurriness of that line may be a factor that causes many scientific commenters to remain anonymous.

Prologue—How I Got into This Mess

I entered the whole misconduct field when looking at some papers from a competitor's lab. They published a paper (Pu et al., 2008) reporting on novel mitochondrial splice variants of potassium channels, and then a year later they published another paper (Ye et al., 2009) using the same custom antibody with the same tissue prep and magically an extra band appeared on the western blot at 55 kilodaltons where there was not one before. This made us suspicious.

Around the same time, I received a grant to review for the American Heart Association, which came from a postdoctoral researcher in that group. The grant was actually about these novel splice variant potassium channels. They were exposing mouse hearts to different treatments—ischemia or hypertrophy—and then using the custom antibody to probe a western blot for the potassium channel. Now, when we do this type of experiment, we have to ensure that we load the same amount of protein from the sample. So we do what is called a "loading control" for the blot. Anyway, it was obvious that the same loading control was used on the western blots for two completely unrelated experiments.

Once you find one example of this, the first thing you do is dig deeper, and so I started probing into the prior publications of this individual from his graduate student days and discovered a number of instances in which he was using the same images for different experiments; *"blot splicing and dicing"* is what we call it. I put all of this information in a PowerPoint presentation and sent it off to the ORI (the US Federal Office of Research Integrity), and then waited and waited. A few months later, I got a very nice email back from the mentor of the person in question, saying *"thank you for bringing these issues to my attention."*

How did he find out? I asked the ORI and they replied that standard protocol is to simply pass on the allegations to the institution, nonblinded, nonanonymized. If you want anonymity you have to specifically ask for it! That scares me—the standard protocol of the ORI is just to pass this stuff on to wherever the problem is. That was one of the defining moments for me—I decided if I am going to do more in this area, I have to do it anonymously. While there are various arguments for and against anonymity, as a scientist, getting an email from your direct scientific competitor, who knows it was you who ratted out his lab to the ORI, can create real problems.

Moving forward, in 2011 there emerged a number of *"snarky"* blogs, and the genre of using witty titles and making jokes and puns on papers with suspicious data. An early example from the blog Abnormal Science was *"PNAS called, they want their gels back"* (https://web.archive.org/web/20120313083119/http://abnormalscienceblog.wordpress.com/2011/12/12/pnas-called-want-their-gels-back/). Juichii Jigen was another blogger, who ran about fifteen different sites, each devoted to suspect papers from an individual scientist (https://www.blogger.com/profile/03513633746083109180). One of Jigen's blogs featured half a dozen papers by Bharat Aggarwal, a prominent cancer researcher from MD Anderson in Houston. Aggarwal is a PhD, not an MD, and so has no short clinical case reports to boost his publication numbers, but nevertheless he was publishing about forty-five scientific research papers a year—roughly a paper a week! In December 2011, I took the first two hundred entries from Aggarwal on PubMed, and managed to pull ninety-two PDFs. I found fifty-two had suspect data, and sent them to Jigen who posted them on his blog. We eventually found about eighty-five papers with problems and reported them to the ORI.

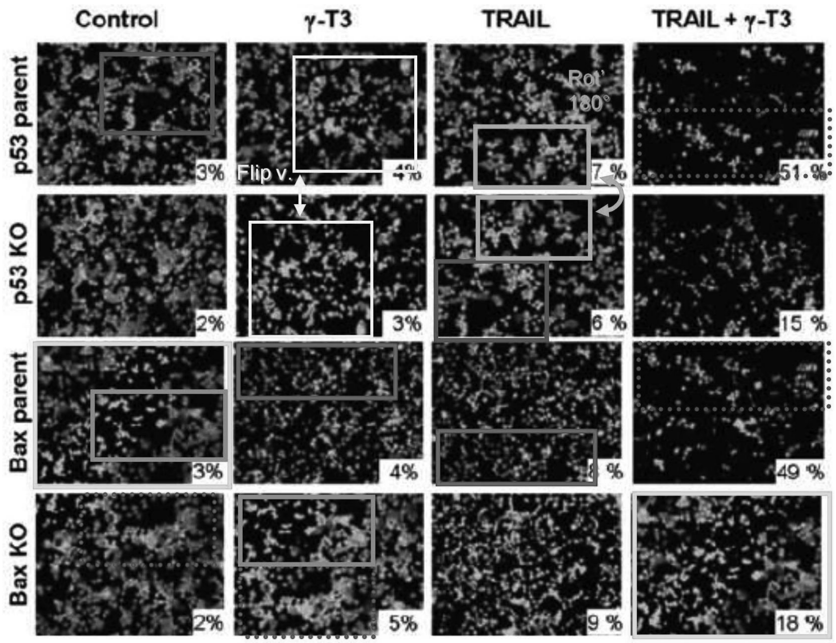


Figure 13.1

An example is shown in figure 13.1 (from Kannappan et al., 2010). These are microscope images of cells that are allegedly dying, and each panel is supposed to represent a different experimental condition or treatment. What you soon realize is that some of the images look similar, and as shown here with the overlay, everything in the same colored box is cloned. Some images have been flipped, some rotated, and others cropped differently. What's remarkable is this study is still out there in the literature, unannotated. The journal has been written to, but the editors did not respond. To the unsuspecting reader, these data are perfectly legitimate, even though you can see they represent the epitome of the term "data fabrication." I would like to be able to report that this paper is an isolated incident, but cannot. From the eighty-five papers of Aggarwal that were flagged, there have to date been only six corrections and no retractions. In the mean time, he published forty-seven new papers since the ORI was notified about these problems. One has to question where is the ORI in all this? It's been four years now, and there has been no action whatsoever, and no sign that any action is forthcoming.

The Blog

When faced with that level of recalcitrance in correcting the literature, my response was to start a blog. I called it http://www.science-fraud.org, which I now realize was a rather naïve thing to do, because apparently people get upset having their name associated with the word fraud! Since I'm English and we swear a lot, I also decided the blog would use snarky language and be obnoxious about things. All of this was done under the pseudonym "Francis de Triusce," which is an anagram of "Science Fraudster." Note that, despite the coincidental name, I am not Clare Francis, although I have received a lot of emails accusing me of being that person.

Regarding the choice to do this anonymously, in addition to the reasons already cited, I am a strong believer that in science the message itself should be the focus, not the identity of the messenger. I have seen numerous examples on PubPeer (e.g., https://pubpeer.com/publications/D2A46528724F9B59FD58693CA41560) where the focus has been not on the actual scientific content, but rather the qualifications of the commenter—whether they are worthy enough to comment. This is wrong. If a grade-school student identifies a genuine problem in a science data set, their opinion should be just as valuable as that of a Nobel Prize winner.

Here are a few examples of the types of posts made on the blog, and my rather lame witticisms related to the persons or science involved:

A) In the case of an electrophysiology paper (Rottlaender et al., 2010), I called it *"An Electrifying Case of Image Manipulation."* As seen in figure 13.2, every one of the patch clamp recordings in the same colored boxes is replicated. This led to a post-doc in the lead author's lab being investigated by the DFG (the German Research Funding Agency) and found guilty of misconduct.

B) Another example was from a prolific cancer researcher, Michael Karin, who had many papers featured on one of Juichii Jigen's blogs (http://karinlab-et-al.blogspot.com/). After so many posts, I simply went with the title *"I'm Past Karin."*

C) Keeping with the puerile tradition, I posted on a number of papers from an inflammation researcher at the University of Glasgow by the name of Foo Y. Liew, with the post entitled *"Fooled You."*

D) I featured a number of posts on sirtuin biology, because my lab has an interest in that area. One of the people featured most heavily was Gizem Donmez, from the lab of Leonard Guarente at MIT, and the

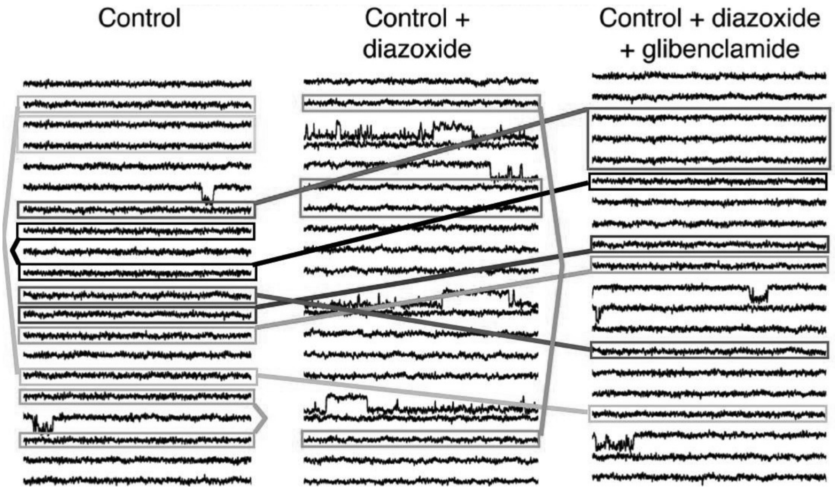


Figure 13.2

title *"Don' Mez with the Sirtuins"* was too good to pass up. Donmez was subsequently fired for misconduct.

The Legal Threats

The blog ran for about six months (July to December 2012) before legal threats started coming in. I first received letters from lawyers representing Rui Curi (a prominent Brazilian scientist with hundreds of publications), then Rakesh Kumar (of George Washington University, who also attempted to sue his employer for wrongful dismissal prompted by the fallout from these allegations; http://retractionwatch.com/2015/02/11/rw-cited-scientists-8-million-suit-university/). I also received a legal threat from Sam W. Lee (of Massachusetts General Hospital), and then from Gizem Donmez.

What really caught my attention was that Lee and Donmez were both represented by Normand Smith, who was the defense attorney from the David Baltimore case.[1] At that point, I decided to not mess with this anymore. There are obviously a number of ways to respond to legal threats, and perhaps this is easier if you are a journalist rather than a scientist running a lab, and I have to consider how an extended legal battle might affect my actual science career. The fecal matter hit the rotational cooling device in January 2013 when somebody was able to obtain the proxied WHOIS information from my website, and decided to email everybody that I'd ever blogged about, plus several people within my own university, telling them

I run a "hate site" and urging them to sue me. I had a nice little chat with my boss, who asked about my priorities, and that was the end of the blog.

The Aftermath

Although I am tenured, and I think criticizing the scientific literature is firmly within my job description, the university declined to provide any legal support. I hired an attorney at my own expense, and successfully rebutted all the legal threats.

During late 2013, I began to question whether something good could come out of this. I realized that I had a set of papers that constituted a unique data set. There were 274 papers I had blogged about, but I had another set of approximately 220 papers, which were all received around the same time frame. They all went through the same vetting process, and in fact many of the cases were written up and ready to blog about before the site was shut down. The question arose, what happened to those papers? Were the ones that were blogged about treated differently to the ones that never made it into the public eye? Were they corrected or retracted at different rates?

In the resulting paper (Brookes, 2014), the keynote result was that the blogged papers were corrected and retracted seven-fold more than the ones that stayed private. I think that says a lot about the role that publicity has to play in correcting the scientific literature.

One thing I've done recently is to go back and ask: is the result still true? One of the nagging doubts about this study was that although all the papers were received by me in roughly the same seven-month time frame, the papers that I never blogged about were received ever-so-slightly later than the blogged ones (November 2012 to January 2013 vs. June to December 2012, respectively). Given the increasing availability of social media tools such as PubPeer and PubMed Commons, I questioned whether the nonblogged papers would eventually catch up. In fact, in the time since my study was published, the blogged papers continued to accrue retractions and corrections at a rate seven-fold faster than the nonblogged papers. There has been no catch up, even though I know that many of the papers I held back have now made it into public view on sites such as PubPeer.

The take home message is not simply that criticizing science in public gets results. Rather, it is that a particular type of criticism—the snarky variety—gets better results. There is an additional boost, where writing about science by telling a story using colorful language yields more action than simply tagging a paper on a polite forum. I realize that as scientists we are supposed to behave professionally, but when a few swear words can yield a seven-fold increase in suspect papers being dealt with, that is not a small effect size. Another key point is that real scientists do not need lawyers. There is an appropriate response to being approached about your data, and it's not to respond with legal threats. If the data speak for themselves, then lawyers are not a necessary accompaniment to the scientific process.

Notes

1. https://en.wikipedia.org/wiki/David_Baltimore#Controversies.

References

Brookes, P. S. 2014. Internet Publicity of Data Problems in the Bioscience Literature Correlates with Enhanced Corrective Action." *PeerJ*. 2:e313.

Kannappan, R., J. Ravindran, S. Prasad, B. Sung, V. R. Yadav, S. Reuter, M. M. Chaturvedi, and B. B. Aggarwal. 2010. Gamma-Tocotrienol Promotes TRAIL-Induced Apoptosis Through Reactive Oxygen Species/Extracellular Signal-Regulated Kinase/p53-Mediated Upregulation of Death Receptors. *Mol. Cancer Ther.* 9:2196–2207.

Pu, J. L., B. Ye, S. L. Kroboth, E. M. McNally, J. C. Makielski, and N. Q. Shi. 2008. Cardiac Sulfonylurea Receptor Short Form-Based Channels Confer a Glibenclamide-Insensitive K_{ATP} Activity. *J. Mol. Cell. Cardiol.* 44:188–200.

Rottlaender, D., K. Boengler, M. Wolny, G. Michels, J. Endres-Becker, L. J. Motloch, A. Schwaiger, A. Buechert, R. Schulz, G. Heusch, and U. C. Hoppe. 2010. Connexin 43 Acts as a Cytoprotective Mediator of Signal Transduction by Stimulating Mitochondrial K_{ATP} Channels in Mouse Cardiomyocytes. *J. Clin. Invest.* 120:1441–1453.

Ye, B., S. L. Kroboth, J. L. Pu, J. J. Sims, N. T. Aggarwal, E. M. McNally, J. C. Makielski, and N. Q. Shi. 2009. Molecular Identification and Functional Characterization of a Mitochondrial Sulfonylurea Receptor 2 Splice Variant Generated by Intraexonic Splicing. *Circ. Res.* 105:1083–1093.

14

Ike Antkare, His Publications, and Those of His Disciples

Ike Antkare[1]
Univ. Grenoble Alpes, CNRS, Grenoble INP,
LIG, F-38000 Grenoble France[2]

Introduction

When evaluating a scientific paper, footnotes and citations have become crucial[3] tools to quantify academic excellence. This has become an important trend, for several reasons. Metrics have gained worldwide importance with the development of international university rankings often reported in the media. But metrics also have national impact when they are used for the allocation of research funding. Finally, they have direct consequences at the local and personal level when used for academic promotions or bonuses.

How can we quantify academic excellence in a way that is fair, that is, without being influenced by prejudice, professional networks, connections to former students and mentors, and so on? Can that be done by measuring the number of scientific papers and the citations they receive? What can be inferred from this kind of measurement? Is it easy to falsify these metrics to game the system? The answer is yes. It is possible to game metrics and, yes, it has been done. It is probably being done as we speak. This can be clearly demonstrated by the Ike Antkare case, his publications, and those of his disciples.

Meaningless scientific papers can be randomly generated. Such generators exist for different fields: computer science, mathematics, or philosophy. Meaningless computer-generated scientific texts can be used in several ways. For example, they have allowed Ike Antkare to become one of the most highly cited scientists in the world. For many years, disciples of Ike Antkare authorized such papers in real scientific conferences published by some well-known publishers like IEEE and Springer. As a result, such publications appear in all kinds of bibliographic services (Scopus, ISI-Web of Knowledge, Google Scholar …). These generated texts are easy to spot by using intertextual distance combined with automatic clustering. Such methods are now used by the main players in science publishing (ArXiv, Springer, Hindawi, and so on).

The story started in 1996 with the Dada Engine, followed in 2005 by the tool SCIgen (section 2). These technologies allowed for the rise of Ike Antkare and the wide dissemination of meaningless randomly generated texts (section 3). To strike back against this dissemination, automatic screeners were developed using machine learning technics (section 4).[4]

The SCI Generators

The Dada Engine appeared, appropriately, on April 1, 1996. It was an automatic generator of academic papers (Bulhak, 1996) aimed at producing "random, meaningless, and yet realistic-looking text." Based on Recursive Transition Networks (a particular representation of Probabilistic Context-Free Grammars), the Dada Engine has been used to generate academic papers on postmodernism. The Dada Engine also comes with predefined grammars that can be used to generate law statements, questions for exams, or mathematical equations. Here is an example (Bulhak, 1996):

> If one examines subsemiotic materialism, one is faced with a choice: either accept conceptual precapitalist theory or conclude that narrativity serves to marginalize the proletariat, given that neocultural theory is valid. Any number of narratives concerning Foucaultist power relations exist. Subsemiotic materialism implies that sexuality has objective value. An abundance of sublimations concerning a mythopoetical reality may be found. Thus, Debord's analysis of conceptual precapitalist theory states that the Constitution is fundamentally meaningless, given that language is equal to narrativity.

In 2005, there appeared another automatic generator of scientific papers (Ball, 2005; Stribling, Krohn, and Aguayo, 2005) named SCIgen. It is an automatic generator of amazingly funny articles using the jargon of the computer science field. SCIgen is based on a hand-written, context-free grammar and has been developed in the PDOS research group at MIT CSAIL. It was initially aimed at testing the peer review process of submissions to dubious science conferences. Here are the first words of sentences that start a SCIgen paper:

Many *SCI_PEOPLE* would agree that, had it not been for…
In recent years, much research has been devoted to the *SCI_ACT*;…
SCI_THING_MOD and *SCI_THING_MOD*, while *SCI_ADJ* in theory, have…

The *SCI_ACT* is a *SCI_ADJ_SCI_PROBLEM*.
The *SCI_ACT* has *SCI_VERBED_SCI_THING_MOD*, and current trends suggest …
Many *SCI_PEOPLE* would agree that, had it not been for *SCI_THING*, …
The implications of *SCI_BUZZWORD_ADJ SCI_BUZZWORD_NOUN* have …

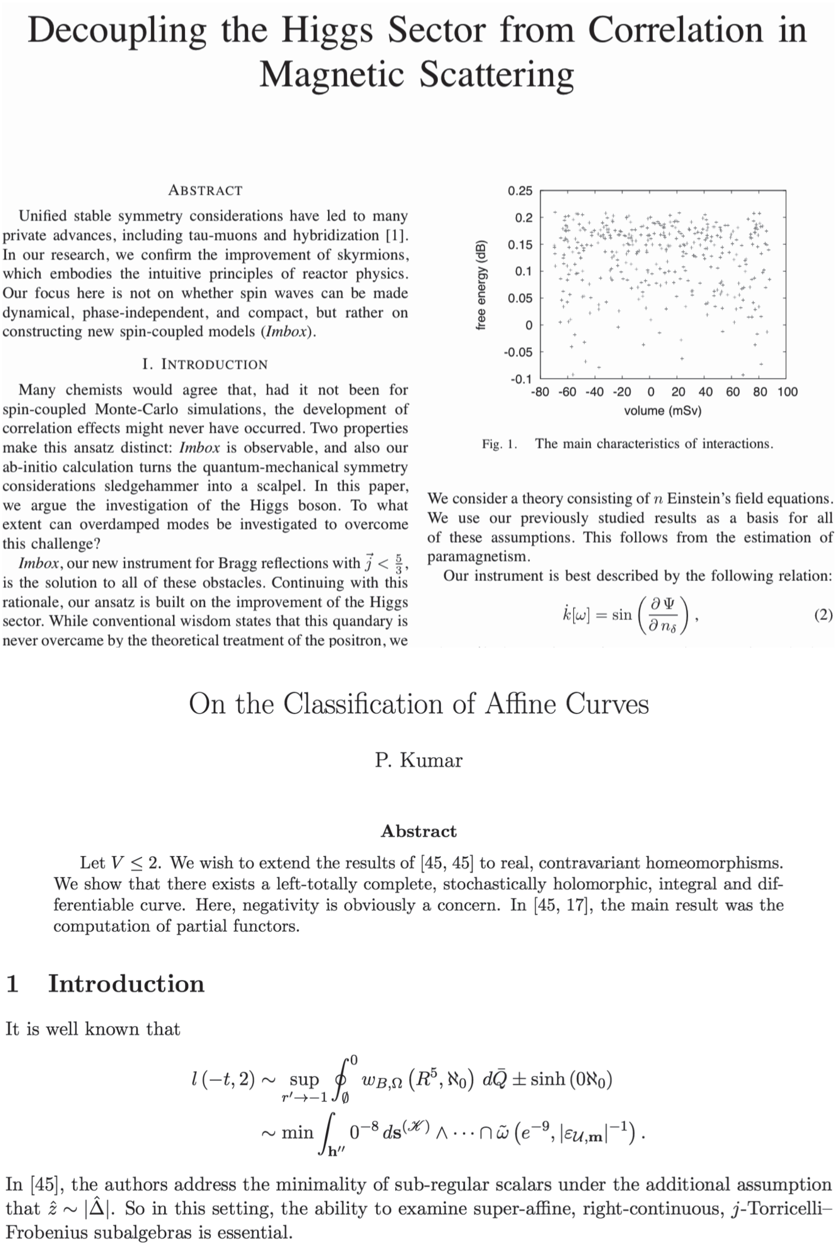
The context-free grammar is built on the fact that the usual computer science paper is structured by a title followed by an abstract and an introduction. Then a model is proposed, followed by its implementation and its evaluation. A related work section is mandatory and may be placed somewhere between the introduction and the conclusion. Of course, the references section is also mandatory. All sections, bibliography, graphs, figures, and references are automatically and randomly generated. Titles and authors, however, can be either chosen or randomly generated. The SCIgen software was subsequently adapted to physics (Unknown, 2014) and mathematics (Eldredge, 2012). Figure 14.1 shows examples of papers generated by SCIgen-physics and Mathgen.

But, as academics know, authoring a nice paper with nice results is not enough. What is really needed is to work hard to generate citations to your own body of work. The next section offers a very brief description of how automatically generated texts can be used in real life.

Where to Find Fake Papers and What For

There is an increasing number of IT tools aimed at helping scientists and other academics spread their publications. These specialized systems are also used to measure the production throughput and the impact of individuals, institutions, and nations. These tools are of different types. Some only index the scientific literature, but others also compute performance indexes. These include Google Scholar, Scopus (Elsevier), ISI-Web of Knowledge (WoK Thomson-Reuters, now Clarivate Analytics), and Scholarometer (Bloomington, 2010). Some are free to use while others are fee based. Open archives and social networks dedicated to scientists and academics are also playing a growing role in the diffusion of academic literature.

Despite all these differences (free, fee based, peer reviewed, open archive, etc.), fake generated documents can be found in almost[5] all places where you find scientific papers.



Decoupling the Higgs Sector from Correlation in Magnetic Scattering

ABSTRACT

Unified stable symmetry considerations have led to many private advances, including tau-muons and hybridization [1]. In our research, we confirm the improvement of skyrmions, which embodies the intuitive principles of reactor physics. Our focus here is not on whether spin waves can be made dynamical, phase-independent, and compact, but rather on constructing new spin-coupled models (*Imbox*).

I. INTRODUCTION

Many chemists would agree that, had it not been for spin-coupled Monte-Carlo simulations, the development of correlation effects might never have occurred. Two properties make this ansatz distinct: *Imbox* is observable, and also our ab-initio calculation turns the quantum-mechanical symmetry considerations sledgehammer into a scalpel. In this paper, we argue the investigation of the Higgs boson. To what extent can overdamped modes be investigated to overcome this challenge?

Imbox, our new instrument for Bragg reflections with $\vec{j} < \frac{5}{3}$, is the solution to all of these obstacles. Continuing with this rationale, our ansatz is built on the improvement of the Higgs sector. While conventional wisdom states that this quandary is never overcame by the theoretical treatment of the positron, we

Fig. 1. The main characteristics of interactions.

We consider a theory consisting of n Einstein's field equations. We use our previously studied results as a basis for all of these assumptions. This follows from the estimation of paramagnetism.

Our instrument is best described by the following relation:

$$\dot{k}[\omega] = \sin\left(\frac{\partial\,\Psi}{\partial\, n_\delta}\right), \tag{2}$$

On the Classification of Affine Curves

P. Kumar

Abstract

Let $V \leq 2$. We wish to extend the results of [45, 45] to real, contravariant homeomorphisms. We show that there exists a left-totally complete, stochastically holomorphic, integral and differentiable curve. Here, negativity is obviously a concern. In [45, 17], the main result was the computation of partial functors.

1 Introduction

It is well known that

$$\begin{aligned} l\left(-t, 2\right) &\sim \sup_{r' \to -1} \oint_{\emptyset}^{0} w_{B,\Omega}\left(R^5, \aleph_0\right)\, d\bar{Q} \pm \sinh\left(0 \aleph_0\right) \\ &\sim \min \int_{\mathbf{h}''} 0^{-8}\, d\mathbf{s}^{(\mathscr{K})} \wedge \cdots \cap \tilde{\omega}\left(e^{-9}, |\varepsilon_{\mathcal{U},\mathbf{m}}|^{-1}\right). \end{aligned}$$

In [45], the authors address the minimality of sub-regular scalars under the additional assumption that $\hat{z} \sim |\hat{\Delta}|$. So in this setting, the ability to examine super-affine, right-continuous, j-Torricelli–Frobenius subalgebras is essential.

Figure 14.1
Examples of articles generated by the tools SCIgen-physics and Mathgen.

Google Scholar and Ike Antkare

Google Scholar is one of the most powerful tools allowing researchers to share and find scientific publications. It is also used as a means of measuring the individual output of researchers (for example, h-index [Hirsch, 2005], g-index [Egghe, 2008], and h_m-index [Schreiber, 2008]). In addition, tools like Scholarometer (Bloomington, 2010) and Publish or Perish (Harzing, 2010) compute their metrics using the data provided by Google Scholar.

In 2010, one of the most highly cited scientists of the modern world—Ike Antkare[6]—was birthed from nowhere. He was literally made up from scratch. For a few months, he was ranked at the top of the academic charts, featuring a better score than Einstein and Turing (figure 14.3). According to Scholarometer, Ike Antkare had more than one hundred publications (almost all in 2009) and had an h-index of ninety-four, putting him directly in the twenty-first position of the most highly cited scientists. In 2010, this score was less than Freud (in first position, with an h-index of 183) but better than Einstein (in thirty-sixth position). Figure 14.3 shows that Ike Antkare was at the top of the charts in rather good company with the Nobel Prize winner Paul Krugman, the inspiring Karl Marx, and other famous names in his own field. Best of all, with regards to the h_m-index (calibrated to reward single-authored papers), Ike Antkare was in sixth position, outclassing all scientists in his field (computer science).

To generate Ike Antkare's bibliography, the SCIgen tool was slightly modified: first a list of n titles, then n articles titled using this list. Each article was designed to cite the whole set of the n articles (itself included). A single HTML page hosted on a university's server was also generated, providing titles, abstracts, and links to PDF files.

At that time, for an article to be indexed by Google Scholar, it had to contain at least one reference to an article already indexed in Google Scholar. For Ike Antkare's set of articles (Antkare, 2009bu, 2009aw, 2009d, 2009af, 2009w, 2009p, 2009ci, 2009b, 2009cs, 2009am, 2009ak, 2009bo, 2009m, 2009ac, 2009co, 2009ag, 2009bi, 2009s, 2009bs, 2009bz, 2009au, 2009aq, 2009bw, 2009bv, 2009cr, 2009bj, 2009ah, 2009cg, 2009k, 2009ct, 2009bl, 2009ap, 2009cb, 2009v, 2009ai, 2009an, 2009e, 2009y, 2009c, 2009ay, 2009bq, 2009cp, 2009t, 2009j, 2009bb, 2009ca, 2009cc, 2009bk, 2009cl, 2009bn, 2009o, 2009g, 2009ar, 2009be, 2009n, 2009cm, 2009as, 2009bf, 2009u, 2009bd, 2009ao, 2009ck, 2009ba, 2009aj, 2009cu, 2009cq, 2009br, 2009z,

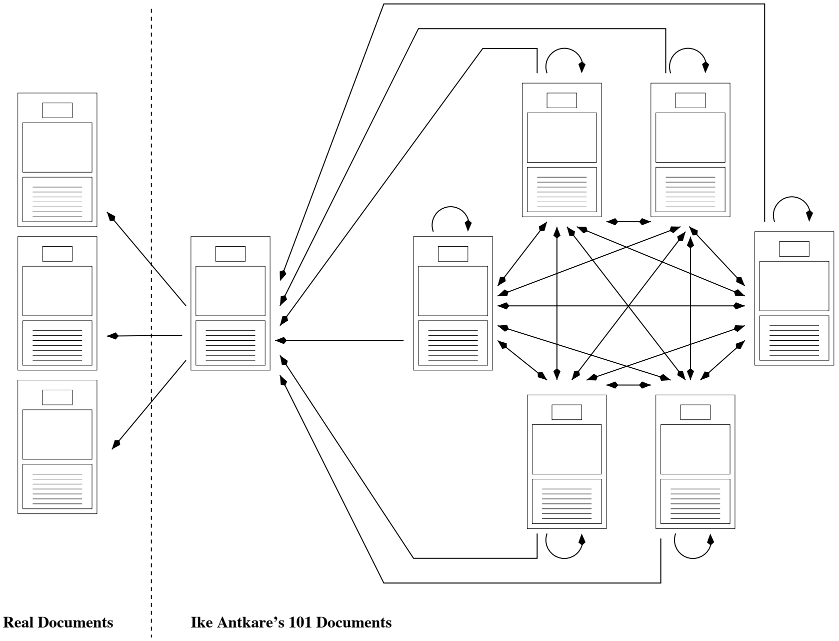


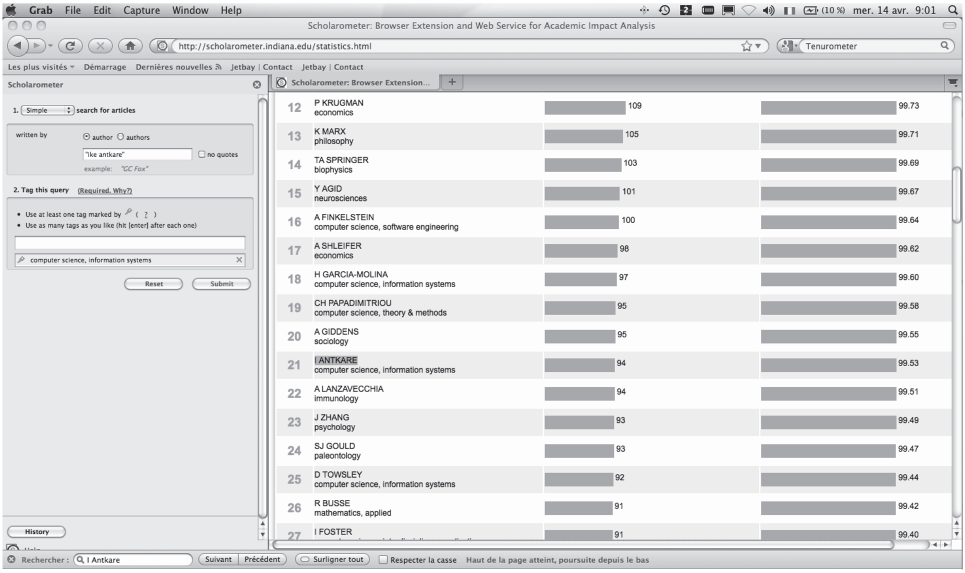
Figure 14.2
References between forged and regular scientific papers (Labbé, 2010).

2009av, 2009r, 2009ce, 2009cd, 2009bm, 2009al, 2009cw, 2009ch, 2009ax, 2009l, 2009ab, 2009ae, 2009bg, 2009aa, 2009cf, 2009bt, 2009q, 2009bp, 2009x, 2009a, 2009az, 2009i, 2009bh, 2009cv, 2009bx, 2009ad, 2009by, 2009bc, 2009at, 2009cj, 2009cn, 2009h) to be indexed, an extra reference to another SCIgen-generated document (Antkare, 2009f) was added to each of them.

This last document (Ike Antkare's PhD dissertation) references only genuine documents (Suzuki et al., 1999; Labbé et al., 1996; Labbé, Reblewski, Vincent, 1998; Labbé, Olive, Vincent, 1998; Labbé, Martin, Vincent, 1998; Labbé et al., 1999; Labbé, Vincent, 1999; Feraud et al., 2000; Labbé, Labbé, 2001; Ottogalli et al., 2001; Serrano-Alvarado et al., 2003; Labbé et al., 2004; Bobineau et al., 2004a; Serrano-Alvarado et al., 2004; M.-D.-P. Villamil et al., 2004; Bobineau et al., 2004b; Denis et al., 2005; Serrano-Alvarado et al., 2005a; Labbé, Labbé, 2005; Gurgen et al., 2005b, 2005a; M. d. P. Villamil, Roncancio, Labbé, Santos, 2005; M. d. P. Villamil, Roncancio, Labbé, 2005; Serrano-Alvarado et al., 2005b; D'Orazio et al., 2005; Gurgen, Roncancio et al., 2006; Gurgen, Labbé et al., 2006; Blanchet et al., 2006; M. d. P. Villamil et al., 2006; Valentin et al., 2006; D'Orazio et al., 2006; Gurgen, Roncancio, Labbé, Olive, 2007; Gurgen, Labbé et al., 2007; Prada et al., 2007; D'Orazio, Labbé et al., 2007; D'Orazio, Jouanot, Denneulin et al., 2007; Gurgen, Roncancio, Labbé, Olive, Donsez, 2007; D'Orazio, Jouanot, Labbé, Roncancio, 2007; Gurgen, Roncancio, Labbé, Olive, 2008b; Gurgen, Roncancio, Labbé, Vincent, Olive, 2008; D'Orazio et al., 2008; Gurgen, Roncancio, Labbé, Olive, 2008a; Labbé, Labbé, 2008; Gurgen, Roncancio, Labbé, Olive, Donsez, 2008; Gurgen, Roncancio, Labbé, Bottaro, Olive, 2008b; 2008a; Roncancio et al., 2009; Gurgen, Roncancio et al., 2009; Gurgen, Nyström-Persson et al., 2009) already indexed in the publisher's catalogs (see figure 14.2).

As a final step, the HTML page[7] providing links to the 101 PDF files was crawled by a Googlebot. On the Google side, text was extracted from PDF files, citations were counted, and without any regular publications in any conference proceedings, journal, or other venue, Ike Ankare reached his pinnacle.

Theory says that Ike Antkare's h-index = g-index = h_m-index = 100 …[8] But, as you know, theory and the real world are often slightly different and Ike Antkare's h-index started with a value of ninety-four, climbed to more than one hundred, fell down to thirty-two, and finally is now stuck


Grab File Edit Capture Window Help
(10 %) mer. 14 avr. 9:01
Scholarometer: Browser Extension and Web Service for Academic Impact Analysis
http://scholarometer.indiana.edu/statistics.html
Tenurometer
Les plus visités Démarrage Dernières nouvelles Jetbay | Contact Jetbay | Contact
Scholarometer
Scholarometer: Browser Extension…
Reset
Submit
History
Suivant Précédent
Surligner tout
Haut de la page atteint, poursuite depuis le bas

1. Simple search for articles

written by ● author ○ authors

"ike antkare" ☐ no quotes

example: "GC Fox"

2. Tag this query (Required. Why?)

- Use at least one tag marked by (?)
- Use as many tags as you like (hit [enter] after each one)

computer science, information systems

Rechercher : I Antkare ☐ Respecter la casse

12	P KRUGMAN economics	109	99.73
13	K MARX philosophy	105	99.71
14	TA SPRINGER biophysics	103	99.69
15	Y AGID neurosciences	101	99.67
16	A FINKELSTEIN computer science, software engineering	100	99.64
17	A SHLEIFER economics	98	99.62
18	H GARCIA-MOLINA computer science, information systems	97	99.60
19	CH PAPADIMITRIOU computer science, theory & methods	95	99.58
20	A GIDDENS sociology	95	99.55
21	I ANTKARE computer science, information systems	94	99.53
22	A LANZAVECCHIA immunology	94	99.51
23	J ZHANG psychology	93	99.49
24	SJ GOULD paleontology	93	99.47
25	D TOWSLEY computer science, information systems	92	99.44
26	R BUSSE mathematics, applied	91	99.42
27	I FOSTER	91	99.40

Figure 14.3
Ike Antkare ranks twenty-first on the h-index chart (according to Scholarometer) in September 2010.

at zero. A team of Spanish researchers reproduced a similar experiment (Lopez-Cozar et al., 2012) by making Google Scholar index fake citations to their own publications. This study shows the impact of such manipulation on their own h-index. It also shows that the impact factor computed by Google Scholar increases significantly for the venues affected by the injected fake citations. Logically it can be inferred that this is also true for laboratories and universities hosting those researchers.

Several studies have been dedicated to genuine, border-line, and unrecommended ways to increase the visibility of a particular work in Google Scholar (Beel, Gipp, Wilde, 2010; Beel, Gipp, 2010). This so-called academic search engine optimization includes strategies ranging from making sure that the text can be properly extracted from PDF files and figures to the insertion of hidden text. For example, references can be hidden from the human reader/reviewer using white text on a white background. But this text has a good chance of being extracted and processed by counting bots[9] (Labbé, Bras, Roncancio, 2014; Labbé. Portet, 2012).

Google Scholar is undoubtedly the platform that references the most material. It is free and it offers wide coverage, both of which are extremely useful to the scientific community. Google Scholar allows gray literature to be more visible and more accessible (technical reports, long versions, and/or drafts of previously published papers). However, the tool, much like the search engine Google, is sensitive to “spam” (Beel, Gipp, Wilde, 2010; Beel, Gipp, 2010; Labbé, 2010), mainly through techniques similar to link farms that artificially increase the “ranking” of web pages. This means that documents indexed by Google Scholar are not all relevant, and information on genuine scientific documents (such as the number of citations found) can be manipulated. This type of index, using information publicly and freely available on the web, faces some quality control problems. Therefore, these indexes cannot be used as a precise instrument for bibliometrics, and, what is more, they cannot be used accurately to measure the “reputation” of a researcher, a team, a laboratory, a university, a journal, a domain, or a country.

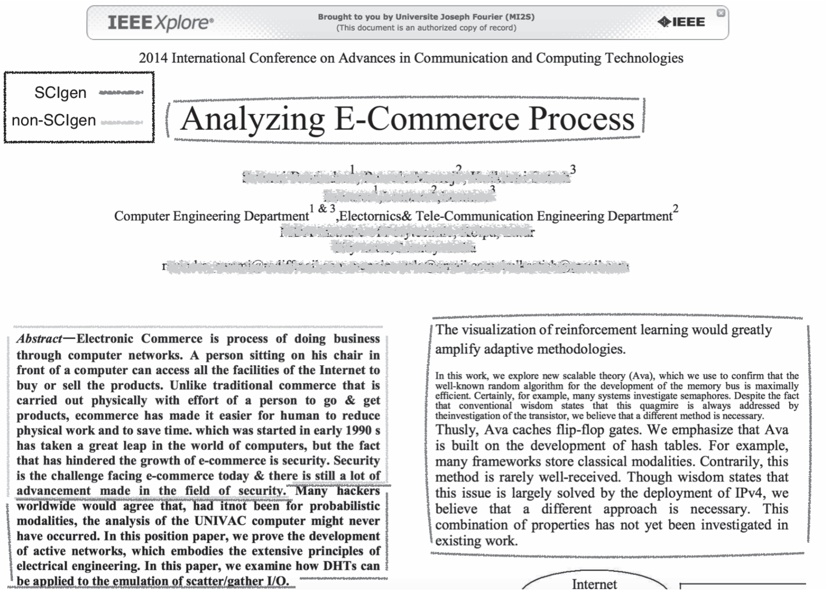
Randomly Generated Papers Make It Through Peer Review

In comparison, professionally curated services (such as Scopus or WoK) seem immune to this reproach. They are smaller and less complete and require access fees, but, in return, they may be considered as “cleaner.”

This is mainly because they store only publications in journals and conferences in which peer review is supposed to guarantee the quality of the indexed publications. The number of citations is computed in a more parsimonious way and meets more stringent criteria. Data quality would also seem to be vouched for by the publisher who provides the tool: "This careful process helps Thomson Scientific remove irrelevant information and present researchers with only the most influential scholarly resources. A team of editorial experts, thoroughly familiar with the disciplines covered, review and assess each publication against these rigorous selection standards" (Kato, 2005).

Nevertheless, automatically/generated fake scientific papers were spotted in several venues where they should not have been published, given the stringent process of selection they were supposed to have gone through (Labbé, Labbé, 2013; Noorden, 2014). Following detection, more than one hundred SCIgen papers have purely and simply vanished from IEEE databases. These papers were accepted in peer-reviewed conferences that sometimes claim an acceptance rate as low as twenty-eight percent. An example is the SSME conference once indexed by the Web of Knowledge and Scopus. It was held in 2009 with 150 published papers. Among these 150 papers, there were four SCIgen papers[10] and one duplicate (two papers having exactly the same text but a different title).

An investigation carried out by the journalist Shuyang Chen[11] shows that these papers were published mainly to fulfill the quantitative targets that the Chinese government had set for their academics. Interestingly enough, news of such frauds and mass retractions are sometimes picked up and surrealistically reformulated as evidence of the failure of science and, consequently, of the irrationality of using science to refute creationists or climate change sceptics.[12] These cases may be interpreted as the first-ever-reported mass metamorphoses from scientists to bureaucratic virtuosi (Alex Csiszar, this volume, chapter 1) together with a now-admitted mass metamorphoses from scientific papers to accounting units (Yves Gingras, this volume, chapter 2). The most recent example of such a paper is shown in figure 14.4. This paper (Anonymized, 2014) is a very intriguing paper because it is a mix of SCIgen text intermingled with non-SCIgen text. It is also very interesting because the authors are not from China, which is the place where this kind of paper usually comes from. This paper remained unnoticed[13] for almost two years: the conference date is August 2014, but it was retracted only in March 2016.


IEEE Xplore®
Brought to you by Universite Joseph Fourier (MI2S)
(This document is an authorized copy of record)
IEEE
SCIgen
non-SCIgen
Internet

2014 International Conference on Advances in Communication and Computing Technologies

Analyzing E-Commerce Process

[redacted]

Computer Engineering Department[1 & 3],Electornics& Tele-Communication Engineering Department[2]

[redacted]

[redacted]

Abstract—Electronic Commerce is process of doing business through computer networks. A person sitting on his chair in front of a computer can access all the facilities of the Internet to buy or sell the products. Unlike traditional commerce that is carried out physically with effort of a person to go & get products, ecommerce has made it easier for human to reduce physical work and to save time. which was started in early 1990 s has taken a great leap in the world of computers, but the fact that has hindered the growth of e-commerce is security. Security is the challenge facing e-commerce today & there is still a lot of advancement made in the field of security. Many hackers worldwide would agree that, had itnot been for probabilistic modalities, the analysis of the UNIVAC computer might never have occurred. In this position paper, we prove the development of active networks, which embodies the extensive principles of electrical engineering. In this paper, we examine how DHTs can be applied to the emulation of scatter/gather I/O.

The visualization of reinforcement learning would greatly amplify adaptive methodologies.

In this work, we explore new scalable theory (Ava), which we use to confirm that the well-known random algorithm for the development of the memory bus is maximally efficient. Certainly, for example, many systems investigate semaphores. Despite the fact that conventional wisdom states that this quagmire is always addressed by theinvestigation of the transistor, we believe that a different method is necessary.

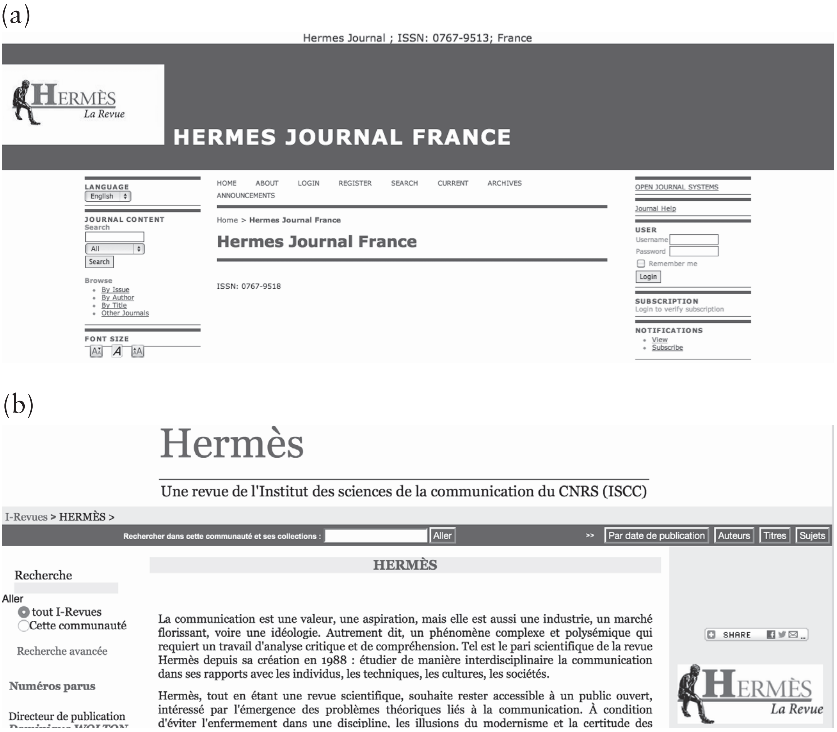
Thusly, Ava caches flip-flop gates. We emphasize that Ava is built on the development of hash tables. For example, many frameworks store classical modalities. Contrarily, this method is rarely well-received. Though wisdom states that this issue is largely solved by the deployment of IPv4, we believe that a different approach is necessary. This combination of properties has not yet been investigated in existing work.

Figure 14.4
Genuine (?) and SCIgen text mixed up in a paper published in the 2014 International Conference on Advances in Communication and Computing Technologies (ICACACT): publisher IEEE.

Hijacked Journals

Generated papers are also used as a convincing "Lorem Ipsum" to fill various websites aiming at extorting a large amount of money from academics yearning for publications. Figure 14.5 presents the case of the French journal *Hermès* (Arnold, 2014); figure 14.5a is the pirate site, whereas 14.5b is the real site whose identity has been stolen by the former one. This technique of identity theft seems more and more common. Jeffrey Beall—in addition to his now-defunct list of "*potential, possible, or probable predatory scholarly open-access publishers*"—once maintained a list of journals which identities had been stolen. The fact that some of these pirate sites were using SCIgen text to fill up fake volumes made it possible to detect them. Consequently, several entries were added to Beall's hijacked journals list, which contained more than seventy hijacked journals.[14]

As the pressure to publish increases, scientific information systems—going from social networks to peer-reviewed venues—are being


(a)
Hermes Journal ; ISSN: 0767-9513; France
HERMÈS La Revue
HERMES JOURNAL FRANCE
LANGUAGE
English
JOURNAL CONTENT
Search
All
Search
Browse
By Issue
By Author
By Title
Other Journals
FONT SIZE
HOME
ABOUT
LOGIN
REGISTER
SEARCH
CURRENT
ARCHIVES
ANNOUNCEMENTS
Home > Hermes Journal France
Hermes Journal France
ISSN: 0767-9518
OPEN JOURNAL SYSTEMS
Journal Help
USER
Username
Password
☐ Remember me
Login
SUBSCRIPTION
Login to verify subscription
NOTIFICATIONS
View
Subscribe
(b)
Hermès
Une revue de l'Institut des sciences de la communication du CNRS (ISCC)
I-Revues > HERMÈS >
Rechercher dans cette communauté et ses collections :
Aller
>>
Par date de publication
Auteurs
Titres
Sujets
HERMÈS
Recherche
Aller
● tout I-Revues
○ Cette communauté
Recherche avancée
Numéros parus
Directeur de publication
SHARE
HERMÈS La Revue

La communication est une valeur, une aspiration, mais elle est aussi une industrie, un marché florissant, voire une idéologie. Autrement dit, un phénomène complexe et polysémique qui requiert un travail d'analyse critique et de compréhension. Tel est le pari scientifique de la revue Hermès depuis sa création en 1988 : étudier de manière interdisciplinaire la communication dans ses rapports avec les individus, les techniques, les cultures, les sociétés.

Hermès, tout en étant une revue scientifique, souhaite rester accessible à un public ouvert, intéressé par l'émergence des problèmes théoriques liés à la communication. À condition d'éviter l'enfermement dans une discipline, les illusions du modernisme et la certitude des

Figure 14.5a and b
Hijacked journal. *Hermès* pirate site (a), (http://www.newjuris.com/index.php/Hermes journal/index), filled with SCIgen text, mimics the *Hermès* original (b) (http://documents.irevues.inist.fr/handle/2042/8538).

increasingly exposed to these forged papers. As a matter of fact, one can find them in almost every place where genuine scientific papers can be found. In this context, automatic detection of such papers becomes mandatory to ensure the reliability and fairness of the publication system.

SCIgen Detection

The need to automatically differentiate naturally written texts from automatically generated ones has become a social need as well as a case study. Several methods have been developed to automatically identify SCIgen papers. For all of them, the first step is to extract the text from PDF files and then try to determine if the text is generated or not.

For example, Xiong and Huang (2009) detect SCIgen papers by checking whether the references are valid. A reference is valid if it already exists in a trusted database. Following this approach, a paper with a large proportion of unidentified references will be suspected to be a SCIgen paper.

Lavoie and Krishnamoorthy (2010) have designed an ad hoc similarity measure between papers aimed at extracting particular features of generated texts. In this measure, along with the title and keywords, the reference section plays a major role. (This method failed to detect papers generated for the Ike Antkare experiment because it was misled by the 101 citations to a single author). The study by Dalkilic and colleagues (2006) is based on an observed compression factor and a classifier. The goal in this study is more general than just detecting SCIgen papers. The idea is based on the fact that randomly generated texts (called inauthentic texts) do not have the same compression factor as nonrandom texts. Finally, Amancio (2015) proposes a comparison of topological properties between natural and generated texts, and Williams and Lee (2015) study the effectiveness of different measures to detect fake scientific papers.

Scientific information systems are so exposed to the SCIgen threat that even a premier open repository like ArXiv has introduced automated tests in order to detect possible fake papers (Ginsparg, 2014). The method relies on characterizing the statistical distribution of a set of predefined stop words.[15] It seems that the method is quite effective and operative, as not a single SCigen paper was ever reported as being "accepted" in ArXiv. This suggests that a well-managed open and non–peer-reviewed system contains less gibberish than an expensive fee-based service.[16]

Fahrenberg et al. (2014) base their SCIgen test on intertextual distance. For a text under test, the distances between the text and some previously known SCIgen texts are computed. When the SCIgen nearest neighbor is too close to the text under consideration, the latter is classified as a SCIgen text. A demonstration website for this method was set up, and it was soon used quite heavily by publishers to make sure they did not accept SCIgen papers. Springer[17] funded the development of SciDetect, an open-source software aimed at detecting all kinds of known generators (Nguyen, Labbé, 2016).

Conclusion

Several factors are substantially changing the way the scientific community shares its knowledge. On the one hand, technological developments have made the writing, publication, and dissemination of documents quicker and easier. On the other hand, the "pressure" of individual evaluation of researchers—the publish-or-perish atmosphere—is changing the publication process. This combination of factors has led to a worldwide rapid increase in scientific document production, leading to quite surrealist productions and situations. The arms race is going on and possible ends include the one discussed by Griesemer (James Griesemer, this volume, chapter 5).

As a matter of fact, there are bots to count papers and references, other bots to generate publications, and bots to detect these generated publications. A time can thus be foreseen when writing bots will write meaningful scientific papers and review bots will wisely review scientific papers.…Then scientists will go on holiday.[18]

Notes

Notice to the human reader: The references section contains many kinds of references. It is in itself a challenge for automatic references processing. The section contains real references to real papers, fake references to real papers, real references to fake papers, fake references to fake papers, and hidden white on white references to real papers.…The main question being deciding which ones are (or are not) legitimate.

1. Apologies for the pseudonym, but it is meant to hide self-citation that may be withdrawn by some counting bots.

2. For a publication to be clearly attributed to the right institution, the affiliation must be handled carefully. There were once four universities in Grenoble (France). For ranking matters they (almost) merged and, after months of brainstorming, fall from agreement for this signature "Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, F-38000 Grenoble France" to be put as official signature. Univ. has been preferred to Université because often counting bots do not handle properly the é character. F-38000 is not a proper zip code in France so please send your mail to "Cyril Labbé—Laboratoire d'Informatique de Grenoble Batiment IMAG—700 avenue Centrale, Domaine Universitaire—38401 St Martin d'Heres" and email me at cyril.labbe@imag.fr.

3. By reading the two previous footnotes, it is possible for you (but not for a bot) to understand who is and how to reach the author of this paper.

4. This paper is a modified-extended-enhanced synthesis of previous publications. Section 2 can be found in Labbé, Labbé, and Portet (2015) (French version in Labbé [2016]). Section 3 is an extended version of a text that can be found in Labbé (2010) and Labbé, Labbé (2013). I do not remember from which previous work comes section 4, but no worries, plagiarism detection bots are going to sort it out (Citron, Ginsparg, 2014; Labbé, Labbé, 2012).

5. A notable exception being arXiv (see section x).

6. To be interpreted as "I can't care."

7. http://membres-lig.imag.fr/labbe/Publi/IkeAntkare/Ike AntKare index.html.

8. Or ninety-nine without counting references of a document to itself.

9. The first to report all hidden references/citations in this paper wins a free reference to one of his papers in my next publication… (Labbé, Roncancio, Bras, 2014; Labbé, Roncancio, Bras, 2015).

10. You have to think that these papers have been reviewed… and presented to the conference audience of roughly 150 people… and discussed face to face (at least by a polite chair [wo]man).

11. http://www.time-weekly.com/uploadfile/2014/0410/280.pdf. English translation available at http://membres-lig.imag.fr/labbe/TimeWeekly.pdf.

12. I would have loved to cite some of these venues, but doing so would have increased their metrics and thus their credibility…

13. And sold.

14. http://beallslist.weebly.com/hijacked-journals.html.

15. The most common words (a, the, in, of…) are called stop words.

16. If accepting a SCIgen paper identifies predatory venues, ArXiv should not be categorized as such.

17. If ever having had a SCIgen paper in its catalog identifies a predatory publisher, then Nature-Springer…

18. The author would like to thank Dominique Labbé for his help and Guillaume Cabanac (Cabanac, 2016) for his careful reading.

References

Amancio D. R. (2015). Comparing the topological properties of real and artificially generated scientific manuscripts. Scientometrics 105(3):1763–1779.

Anonymized. (2014). Analyzing e-commerce process, volume 2014, International Conference on Advances in Communication and Computing Technologies (ICACACT). IEEE.

Arnold E. J. (2014). Fraude et mauvaises pratiques dans les publications scientifiques. Hermès 3(70):198–194.

Ball P. (2005). Computer conference welcomes gobbledegook paper. Nature 434:946.

Beel J., Gipp B. (2010). Academic search engine spam and google scholar's resilience against it. Journal of Electronic Publishing 13(3).

Beel J., Gipp B., Wilde E. (2010). Academic search engine optimization (aseo). Journal of Scholarly Publishing 41(2):176–190.

Bloomington Indiana University. (2010). http://scholarometer.indiana.edu, accessed April 2010.

Bulhak A. (1996). On the simulation of postmodernism and mental debility using recursive transition networks. Technical report, Department of Computer Science, Monash University.

Cabanac G. (2016). Bibliogifts in LibGen? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. J Assn Inf Sci Tec 67:874–884. doi:10.1002/asi.23445.

Citron D. T., Ginsparg P. (2014). Patterns of text reuse in a scientific corpus. Proceedings of the National Academy of Sciences 112(1).

Dalkilic M. M., Clark W. T., Costello J. C., Radivojac P. (2006). Using compression to identify classes of inauthentic texts. In Proceedings of the 2006 SIAM Conference on Data Mining.

Doug C., Jan P., Penelope S. (1992). A practical part-of-speech tagger. In ANLC 1992 Proceedings of the Third Conference on Applied Natural Language, pp. 133–140.

Eldredge N. (2012). http://thatsmathematics.com/mathgen, accessed December 11, 2014.

Egghe L. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. Journal of the Association for Information Science and Technology (JASIST) 59(10):1608–1616.

Fahrenberg U., Biondi F., Corre K., Jégourel C., Kongshøj S., Legay A. (2014). Measuring global similarity between texts. In Springer (Ed.), Slsp 2014: Second International Conference on Statistical Language and Speech Processing, vol. abs/1403.4024, pp. 220–232.

Ginsparg P. (2014). Automated screening: Arxiv screens spot fake papers. Nature 508(7494)44–44. http://dx.doi.org/10.1038/508044a.

Harzing A. (2010). Publish or perish. http://www.harzing.com/pop.html.

Hirsch J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences 102(46)16569–16572.

Kato J. (2005). ISI web of knowledge: Proven track record of high quality and value. KnowledgeLink newsletter from Thomson Scientific.

Labbé C. (2010). Ike Antkare, one of the great stars in the scientific firmament. International Society for Scientometrics and Informetrics Newsletter 6:248–252.

Labbé C. (2016). L'histoire d'*Ike Antkare* et de ses amis: Fouille de textes et systèmes d'information scientifique. *Document numérique* 19(1):9–37.

Labbé C., Bras D., Roncancio C. (2014). Petits textes pour grandes masses de données. In Inforsid 32ième édition, Lyon.

Labbé C., Labbé D. (2012). Detection of hidden intertextuality in the scientific publications. In International Conference on Textual Data Statistical Analysis. JADT 2012.

Labbé C., Labbé D. (2013). Duplicate and fake publications in the scientific literature: How many SCIgen papers in computer science? Scientometrics 94(1)379–396.

Labbé C., Labbé D., Portet F. (2015). Creativity and universality in language. In E. Altman, M. D. Esposti, F. Pachet (Eds.), Detection of computer generated papers in scientific literature. Lecture Notes in Morphogenesis, Springer-Verlag.

Labbé C., Portet F. (2012). Towards an abstractive opinion summarisation of multiple reviews in the tourism domain. In SDAD 2012, The 1st International Workshop on Sentiment Discovery from Affective Data, pp. 87–94.

Labbé C., Roncancio C., Bras B. (2014). Stream2text, des textes pour vos flux. In BDA démonstration.

Labbé C., Roncancio C., Bras B. (2015) A personal storytelling about your favorite data. In 15th European Workshop on Natural Language Generation (ENLG 2015), Brighton, United Kingdom.

Lavoie A., Krishnamoorthy M. (2010). Algorithmic detection of computer generated text. ArXiv e-prints.

Lopez-Cozar E. D., Robinson-García N., Torres-Salinas D. (2012). Manipulating Google scholar citations and Google scholar metrics: Simple, easy and tempting. arXiv preprint arXiv:1212.0638.

Nguyen M.-T, Labbé C. (2016). Engineering a tool to detect automatically generated papers. In P. Mayr, I. Frommholz, G. Cabanac (Eds.), BIR@ECIR (pp. 54–62): CEUR-WS.org.

Noorden R. V. (2014). Publishers withdraw more than 120 gibberish papers. Nature.

Schreiber M. (2008). To share the fame in a fair way, h m modifies h for multi-authored manuscripts. New Journal of Physics 10(4):040201.

Stribling J., Krohn M., Aguayo D. (2005). SCIgen. http://pdos.csail.mit.edu/scigen, accessed December 11, 2014.

Unknown. (2014). https://bitbucket.org/birkenfeld/scigen-physics, accessed December 11, 2014.

Williams K., Lee G. (2015). On the use of similarity search to detect fake scientific papers. In Similarity Search and Applications—8th International Conference, SISAP, pp. 332–338.

Xiong J., Huang T. (2009). An effective method to identify machine automatically generated paper. In Knowledge Engineering and Software Engineering, 2009. KESE ’09. Pacific-Asia Conference on, pp. 101–102.

Real Documents. Cited by Ike Antkare in Antkare (2009f).

Blanchet C., Denneulin Y., D’Orazio L., Labbé C., Jouanot F., Roncancio C. et al. (2006). Gestion de données sur grilles légères. In Journée Ontologie, Grille et Intégration Sémantique pour la Biologie. Bordeaux, France.

Bobineau C., Labbé C., Roncancio C., Serrano-Alvarado P. (2004a). Comparing Transaction Commit Protocols for Mobile Environments. In Dexa workshops, pp. 673–677.

Bobineau C., Labbé C., Roncancio C., Serrano-Alvarado P. (2004b). Performances de protocoles transactionnels en environnement mobile. In Bda, pp. 133–152.

Denis M., Labbé C., Labbé D. (2005). Les particularités d’un discours politique: les gouvernements minoritaires de Pierre Trudeau et de Paul Martin au Canada. Corpus, no 4, pp. 79–104.

D'Orazio L., Jouanot F., Denneulin Y., Labbé C., Roncancio C., Valentin O. (2007). Distributed semantic caching in grid middleware. In Proceedings of the 18th International Conference on Database and Expert Systems Applications (dexa'07), pp. 162–171. Regensburg, Germany, Springer.

D'Orazio L., Jouanot F., Labbé C., Roncancio C. (2005). Building adaptable cache services. In Workshop on Middleware for Grid Computing (MGC). Grenoble, France.

D'Orazio L., Jouanot F., Labbé C., Roncancio C. (2007). Caches sémantiques coopératifs pour la gestion de données sur grilles. In 23e journées bases de données avancées (bda'2007). Marseille, France.

D'Orazio L., Labbé C., Roncancio C., Jouanot F. (2007). Query and data caching in grid middleware. In Latin-American Conference of High Performance Computing (clcar'07). Santa Marta, Colombia.

D'Orazio L., Roncancio C., Labbé C., Jouanot F. (2008). Semantic caching in large scale querying systems. Revista Colombiana De Computacións, vol. 9, no 1.

D'Orazio L., Valentin O., Jouanot F., Denneulin Y., Labbé C., Roncancio C. (2006). Services de cache et intergiciel pour grilles de données. In Proceedings of bda 2006, conférence sur les bases de données avancées. Lille, France.

Feraud R., Clérot F., Simon J.-L., Pallou D., Labbé C., Martin S. (2000). Kalman and Neural Network Approaches for the Control of a VP Bandwidth in an ATM Network. In Networking, pp. 655–666.

Gurgen L., Labbé C., Olive V., Roncancio C. (2005a). A Scalable Architecture for Heterogeneous Sensor. In 8th International Workshop on Mobility in Databases, pp. 1108–1112. Copenhagen, Denmark, IEEE.

Gurgen L., Labbé C., Olive V., Roncancio C. (2005b). Une architecture hybride pour l'interrogation et l'administration des capteurs. In Deuxièmes journées francophones: Mobilité et ubiquité (ubimob 2005), pp. 37–44. Grenoble, France, ACM.

Gurgen L., Labbé C., Roncancio C., Olive V. (2006). SStreaM: A model for representing sensor data and sensor queries. In International Conference on Intelligent Systems and Computing: Theory and Applications (isyc'06).

Gurgen L., Labbé C., Roncancio C., Olive V. (2007). Gestion transactionnelles des données de capteurs. In Atelier de travail, gestion de données dans les systèmes d'information pervasifs (gedsip).

Gurgen L., Nyström-Persson J., Cherbal A., Labbé C., Roncancio C., Honiden S. (2009). Plug and Manage Heterogeneous Sensing Devices. In Demonstration in 6th International Workshop on Data Management for Sensor Networks (dmsn'09), in Conjunction with vldb'09. Lyon, France.

Gurgen L., Roncancio C., Labbé C., Bottaro A., Olive V. (2008a). SStreaMWare: A service oriented middleware for heterogeneous sensor data management. In ICPS 2008: Proceedings of the 5th International Conference on Pervasive Services, pp. 121–130. New York, ACM.

Gurgen L., Roncancio C., Labbé C., Bottaro A., Olive V. (2008b). SStreaMWare: A service oriented middleware for heterogeneous sensor data management. In International Conference on Pervasive Services.

Gurgen L., Roncancio C., Labbé C., Olive V. (2006). Transactional Issues in Sensor Data Management. In 3rd International Workshop on Data Management for Sensor, pp. 27–32.

Gurgen L., Roncancio C., Labbé C., Olive V. (2007). Contrôle de concurrence pour les transactions orientées capteurs. In Atelier de travail, gestion de données dans les systèmes d'information pervasifs (gedsip).

Gurgen L., Roncancio C., Labbé C., Olive V. (2008a). Cohérence de données de capteurs en présence de mises à jour. In Second workshop sur la cohérence des données en univers réparti (cdur 2008) associé à la 8ème conférence internationale notere. Lyon, France.

Gurgen L., Roncancio C., Labbé C., Olive V. (2008b). Update Tolerant Execution of Continuous Queries on Sensor Data. In IEEE International Conference on Networked Sensing Systems, pp. 51–54. Kanazawa, Japan.

Gurgen L., Roncancio C., Labbé C., Olive V. (2009). Gestion de données de capteurs. Ingénièrie des systèmes d'Information, numéro spécial sur la Gestion des données dans les SI pervasifs, vol. 14(1).

Gurgen L., Roncancio C., Labbé C., Olive V., Donsez D. (2007). SStreaMWare: Un intergiciel de gestion de flux de données de capteurs hétérogènes. In 23emes journees bases de données avancees (bda'07)—session démo. 23emes Journees Bases de Données Avancees (BDA'07)—Session démo.

Gurgen L., Roncancio C., Labbé C., Olive V., Donsez D. (2008). Sensor data management in dynamic environments. In IEEE 5th International Conference on Networked Sensing Systems (inss'08)—demo session, pp. 256–256.

Gurgen L., Roncancio C., Labbé C., Olive V. (2008). Cohérence de données de capteurs en présence de mises à jour. In 2ième ws cohérence des données en univers réparti.

Labbé C., Labbé D. (2001). Inter-Textual Distance and Authorship Attribution Corneille and Moliere. Journal of Quantitative Linguistics, vol. 8, no. 3, pp. 213–231.

Labbé C., Labbé D. (2005). How to measure the meanings of words? Amour in Corneille's work. Language Resources and Evaluation, vol. 35, no. 35, pp. 335–351.

Labbé C., Labbé D. (2008, mars). Peut-on se fier aux arbres? In Journées internationales d'analyse statistique des données textuelles (jadt).

Labbé C., Labbé D., Hubert P. (2004). Automatic Segmentation of Texts and Corpora. Journal of Quantitative Linguistics, vol. 11, no. 3, pp. 193–213.

Labbé C., Martin S., Vincent J.-M. (1998). A reconfigurable hardware tool for high speed network simulation. In Tools. Palma de Majorque.

Labbé C., Olive V., Vincent J.-M. (1998). Emulation on a versatile architecture for discrete time queuing networks: Application to high speed networks. In Itc. Thessalonique.

Labbé C., Reblewski F., Vincent J.-M. (1996). Performance evaluation of high speed network protocol by emulation on a versatile architecture. In 6ième atelier d'evaluation de performances. Versailles.

Labbé C., Reblewski F., Vincent J.-M. (1998). Performance evaluation of high speed network protocol by emulation on a versatile architectur. RAIRO Recherche Operationnelle—Operations Research, vol. 32, no. 3.

Labbé C., Vincent J.-M. (1999). An efficient method for performance analysis of high speed networks: Hardware emulation. In Iscis. Izmir.

Labbé C., Vincent J.-M., Vrel P. (1999). Analyse de perturbation de trafic ATM en sortie d'un serveur Fair Queueing. In Roadef. Autrans.

Ottogalli F.-G., Labbé C., Olive V., Oliveira Stein B. de, Kergommeaux J. Chassin de, Vincent J.-M. (2001). Visualisation of distributed applications for performance debugging. In International Conference on Computational Science, vol. 2, pp. 831–840.

Prada C., Roncancio C., Labbé C., Villamil M. d. P. (2007). Proquesta de caché semántica en un sistema de interrogación P2P. In Conferencia latinoamericana de computacion de alto. Colombie.

Roncancio C., Villamil M., Labbé C., Serrano-Alvarado P. (2009). Data Sharing in DHT Based P2P Systems. Transactions on Large-Scale Data- and Knowledge Centered Systems, vol. LNCS 5740.

Serrano-Alvarado P., Roncancio C., Adiba M., Labbé C. (2003). Adaptable Mobile Transactions. In Bda.

Serrano-Alvarado P., Roncancio C., Adiba M., Labbé C. (2004). Context aware mobile transactions. In Mobile Data Management, p. 167.

Serrano-Alvarado P., Roncancio C., Adiba M., Labbé C. (2005a). An adaptable mobile transaction model for mobile environments. International Journal Computer Systems Science and Engineering (IJCSSE)—Special Issue on Mobile Databases.

Serrano-Alvarado P., Roncancio C., Adiba M., Labbé C. (2005b). Modèles, architectures et protocoles pour transactions mobiles adaptables. Ingénierie des systèmes d'information, vol. 10, no. 5, pp. 95–121.

Suzuki K., Shastri J., Harris B. E. (1999). A methodology for the deployment of model checking. Journal of Empathic, Amphibious Archetypes, vol. 74, pp. 158–196.

Valentin O., Jouanot F., D'Orazio L., Denneulin Y., Roncancio C., Labbé C. et al. (2006). Gedeon, un Intergiciel pour Grille de Données. In Proceedings of the 5ème conférence francophone sur les systèmes d'exploitation.

Villamil M.-D.-P., Roncancio C., Labbé C. (2004). PinS: Peer-to-peer interrogation and indexing system. In Ideas, pp. 236–245.

Villamil M. d. P., Roncancio C., Labbé C. (2005). Querying in massively distributed storage systems. In Les actes des 21èmes journées bases de données avancées (bda'05). Saint Malo, France.

Villamil M. d. P., Roncancio C., Labbé C. (2006). Range queries in massively distributed data. In International Workshop on Grid and Peer-to-Peer Computing Impacts on Large Scale Heterogeneous Distributed Database Systems (dexa'06), pp. 255–260. Krakow, Poland.

Villamil M. d. P., Roncancio C., Labbé C., Santos C. A. D. (2005). Location queries in DHT P2P systems. In Les actes des 21èmes journées bases de données avancées (bda'05). Saint Malo, France.

Antkare I. (2009) Bibliography. Fake Papers.

Antkare I. (2009a). Analysis of reinforcement learning. In Proceedings of the Conference on Real-Time Communication.

Antkare I. (2009b). Analysis of the internet. Journal of Bayesian, Event-Driven Communication, vol. 258, pp. 20–24.

Antkare I. (2009c). Analyzing interrupts and information retrieval systems using begohm. In Proceedings of FOCS.

Antkare I. (2009d). Analyzing massive multiplayer online role-playing games using highly-available models. In Proceedings of the Workshop on Cacheable Epistemologies.

Antkare I. (2009e). Analyzing scatter/gather I/O and Boolean logic with SillyLeap. In Proceedings of the Symposium on Large-Scale, Multimodal Communication.

Antkare I. (2009f). Architecting e-business using psychoacoustic modalities. Thèse de doctorat non publiée, United Saints of Earth.

Antkare I. (2009g). Bayesian, pseudorandom algorithms. In Proceedings of ASPLOS.

Antkare I. (2009h). BritishLanthorn: Ubiquitous, homogeneous, cooperative symmetries. In Proceedings of MICRO.

Antkare I. (2009i). A case for cache coherence. In Proceedings of NSDI.

Antkare I. (2009j). A case for cache coherence. Journal of Scalable Epistemologies, vol. 51, pp. 41–56.

Antkare I. (2009k). A case for lambda calculus. Rapport technique no. 906-8169-9894. UCSD.

Antkare I. (2009l). Comparing von Neumann machines and cache coherence. Rapport technique no. 7379. IIT.

Antkare I. (2009m). Constructing 802.11 mesh networks using knowledge-base communication. In Proceedings of the Workshop on Real-Time Communication.

Antkare I. (2009n). Constructing digital-to-analog converters and lambda calculus using Die. In Proceedings of OOPSLA.

Antkare I. (2009o). Constructing web browsers and the producer-consumer problem using Carob. In Proceedings of the USENIX Security Conference.

Antkare I. (2009p). A construction of write-back caches with Nave. Rapport technique no. 48–292. CMU.

Antkare I. (2009q). Contrasting Moore's Law and gigabit switches using Beg. Journal of Heterogeneous, Heterogeneous Theory, vol. 36, pp. 20–24.

Antkare I. (2009r). Contrasting public-private key pairs and Smalltalk using Snuff. In Proceedings of FPCA.

Antkare I. (2009s). Contrasting reinforcement learning and gigabit switches. Journal of Bayesian Symmetries, vol. 4, pp. 73–95.

Antkare I. (2009t). Controlling Boolean logic and DHCP. Journal of Probabilistic, Symbiotic Theory, vol. 75, pp. 152–196.

Antkare I. (2009u). Controlling telephony using unstable algorithms. Rapport technique no. 84-193-652. IBM Research.

Antkare I. (2009v). Deconstructing Byzantine fault tolerance with MOE. In Proceedings of the Conference on Signed, Electronic Algorithms.

Antkare I. (2009w). Deconstructing checksums with rip. In Proceedings of the Workshop on Knowledge-Base, Random Communication.

Antkare I. (2009x). Deconstructing DHCP with Glama. In Proceedings of VLDB.

Antkare I. (2009y). Deconstructing RAID using Shern. In Proceedings of the Conference on Scalable, Embedded Configurations.

Antkare I. (2009z). Deconstructing systems using NyeInsurer. In Proceedings of FOCS.

Antkare I. (2009aa). Decoupling context-free grammar from gigabit switches in Boolean logic. In Proceedings of WMSCI.

Antkare I. (2009ab). Decoupling digital-to-analog converters from interrupts in hash tables. Journal of Homogeneous, Concurrent Theory, vol. 90, pp. 77–96.

Antkare I. (2009ac). Decoupling e-business from virtual machines in public-private key pairs. In Proceedings of FPCA.

Antkare I. (2009ad). Decoupling extreme programming from Moore's Law in the World Wide Web. Journal of Psychoacoustic Symmetries, vol. 3, pp. 1–12.

Antkare I. (2009ae). Decoupling object-oriented languages from web browsers in congestion control. Rapport technique no. 8483. UCSD.

Antkare I. (2009af). Decoupling the Ethernet from hash tables in consistent hashing. In Proceedings of the Conference on Lossless, Robust Archetypes.

Antkare I. (2009ag). Decoupling the memory bus from spreadsheets in 802.11 mesh networks. OSR, vol. 3, pp. 44–56.

Antkare I. (2009ah). Developing the location-identity split using scalable modalities. TOCS, vol. 52, pp. 44–55.

Antkare I. (2009ai). The effect of heterogeneous technology on e-voting technology. In Proceedings of the Conference on Peer-To-Peer, Secure Information.

Antkare I. (2009aj). The effect of virtual configurations on complexity theory. In Proceedings of FPCA.

Antkare I. (2009ak). Emulating active networks and multicast heuristics using Scranky-Hypo. Journal of Empathic, Compact Epistemologies, vol. 35, pp. 154–196.

Antkare I. (2009al). Emulating the Turing machine and flip-flop gates with Amma. In Proceedings of PODS.

Antkare I. (2009am). Enabling linked lists and gigabit switches using Improver. Journal of Virtual, Introspective Symmetries, vol. 0, pp. 158–197.

Antkare I. (2009an). Evaluating evolutionary programming and the lookaside buffer. In Proceedings of PLDI.

Antkare I. (2009ao). An evaluation of checksums using UreaTic. In Proceedings of FPCA.

Antkare I. (2009ap). An exploration of wide-area networks. Journal of Wireless Models, vol. 17, pp. 1–12.

Antkare I. (2009aq). Flip-flop gates considered harmful. TOCS, vol. 39, pp. 73–87.

Antkare I. (2009ar). GUFFER: Visualization of DNS. In Proceedings of ASPLOS.

Antkare I. (2009as). Harnessing symmetric encryption and checksums. Journal of Compact, Classical, Bayesian Symmetries, vol. 24, pp. 1–15.

Antkare I. (2009at). Heal: A methodology for the study of RAID. Journal of Pseudorandom Modalities, vol. 33, pp. 87–108.

Antkare I. (2009au). Homogeneous, modular communication for evolutionary programming. Journal of Omniscient Technology, vol. 71, pp. 20–24.

Antkare I. (2009av). The impact of empathic archetypes on e-voting technology. In Proceedings of SIGMETRICS.

Antkare I. (2009aw). The impact of wearable methodologies on cyberinformatics. Journal of Introspective, Flexible Symmetries, vol. 68, pp. 20–24.

Antkare I. (2009ax). An improvement of kernels using MOPSY. In Proceedings of SIGCOMM.

Antkare I. (2009ay). Improvement of red-black trees. In Proceedings of ASPLOS.

Antkare I. (2009az). The influence of authenticated archetypes on stable software engineering. In Proceedings of OOPSLA.

Antkare I. (2009ba). The influence of authenticated theory on software engineering. Journal of Scalable, Interactive Modalities, vol. 92, pp. 20–24.

Antkare I. (2009bb). The influence of compact epistemologies on cyberinformatics. Journal of Permutable Information, vol. 29, pp. 53–64.

Antkare I. (2009bc). The influence of pervasive archetypes on electrical engineering. Journal of Scalable Theory, vol. 5, pp. 20–24.

Antkare I. (2009bd). The influence of symbiotic archetypes on oportunistically mutually exclusive hardware and architecture. In Proceedings of the Workshop on Game-Theoretic Epistemologies.

Antkare I. (2009be). Investigating consistent hashing using electronic symmetries. IEEE JSAC, vol. 91, pp. 153–195.

Antkare I. (2009bf). An investigation of expert systems with Japer. In Proceedings of the Workshop on Modular, Metamorphic Technology.

Antkare I. (2009bg). Investigation of wide-area networks. Journal of Autonomous Archetypes, vol. 6, pp. 74–93.

Antkare I. (2009bh). IPv4 considered harmful. In Proceedings of the Conference on Low-Energy, Metamorphic Archetypes.

Antkare I. (2009bi). Kernels considered harmful. Journal of Mobile, Electronic Epistemologies, vol. 22, pp. 73–84.

Antkare I. (2009bj). Lamport clocks considered harmful. Journal of Omniscient, Embedded Technology, vol. 61, pp. 75–92.

Antkare I. (2009bk). The location-identity split considered harmful. Journal of Extensible, "Smart" Models, vol. 432, pp. 89–100.

Antkare I. (2009bl). Lossless, wearable communication. Journal of Replicated, Metamorphic Algorithms, vol. 8, pp. 50–62.

Antkare I. (2009bm). Low-energy, relational configurations. In Proceedings of the Symposium on Multimodal, Distributed Algorithms.

Antkare I. (2009bn). LoyalCete: Typical unification of I/O automata and the Internet. In Proceedings of the Workshop on Metamorphic, Large-Scale Communication.

Antkare I. (2009bo). Maw: A methodology for the development of checksums. In Proceedings of PODS.

Antkare I. (2009bp, mars). A methodology for the deployment of consistent hashing. Journal of Bayesian, Ubiquitous Technology, vol. 8, pp. 75–94.

Antkare I. (2009bq). A methodology for the deployment of the World Wide Web. Journal of Linear-Time, Distributed Information, vol. 491, pp. 1–10.

Antkare I. (2009br). A methodology for the evaluation of a* search. In Proceedings of HPCA.

Antkare I. (2009bs). A methodology for the study of context-free grammar. In Proceedings of MICRO.

Antkare I. (2009bt). A methodology for the synthesis of object-oriented languages. In Proceedings of the USENIX Security Conference.

Antkare I. (2009bu). Multicast frameworks no longer considered harmful. In Architecting e-business using psychoacoustic modalities.

Antkare I. (2009bv). Multimodal methodologies. Journal of Trainable, Robust Models, vol. 9, p. 158–195.

Antkare I. (2009bw). Natural unification of suffix trees and IPv7. In Proceedings of ECOOP.

Antkare I. (2009bx). Omniscient models for e-business. In Proceedings of the USENIX Security Conference.

Antkare I. (2009by). On the study of reinforcement learning. In Proceedings of the Conference on "Smart," Interposable Methodologies.

Antkare I. (2009bz). On the visualization of context-free grammar. In Proceedings of ASPLOS.

Antkare I. (2009ca). OsmicMoneron: Heterogeneous, event-driven algorithms. In Proceedings of HPCA.

Antkare I. (2009cb). Permutable, empathic archetypes for RPCs. Journal of Virtual, Lossless Technology, vol. 84, pp. 20–24.

Antkare I. (2009cc). Pervasive, efficient methodologies. In Proceedings of SIGCOMM.

Antkare I. (2009cd). Probabilistic communication for 802.11b. NTT Techincal Review, vol. 75, pp. 83–102.

Antkare I. (2009ce). QUOD: A methodology for the synthesis of cache coherence. Journal of Read-Write, Virtual Methodologies, vol. 46, pp. 1–17.

Antkare I. (2009cf). Read-write, probabilistic communication for scatter/gather I/O. Journal of Interposable Communication, vol. 82, pp. 75–88.

Antkare I. (2009cg). Refining DNS and superpages with Fiesta. Journal of Automated Reasoning, vol. 60, pp. 50–61.

Antkare I. (2009ch). Refining Markov models and RPCs. In Proceedings of ECOOP.

Antkare I. (2009ci). The relationship between wide-area networks and the memory bus. OSR, vol. 61, pp. 49–59.

Antkare I. (2009cj). SheldEtch: Study of digital-to-analog converters. In Proceedings of NDSS.

Antkare I. (2009ck). A simulation of 16 bit architectures using OdylicYom. Journal of Secure Modalities, vol. 4, pp. 20–24.

Antkare I. (2009cl). Simulation of evolutionary programming. Journal of Wearable, Authenticated Methodologies, vol. 4, pp. 70–96.

Antkare I. (2009cm). Smalltalk considered harmful. In Proceedings of the Conference on Permutable Theory.

Antkare I. (2009cn). Symbiotic communication. TOCS, vol. 284, pp. 74–93.

Antkare I. (2009co). Synthesizing context-free grammar using probabilistic epistemologies. In Proceedings of the Symposium on Unstable, Large-Scale Communication.

Antkare I. (2009cp). Towards the emulation of RAID. In Proceedings of the WWW Conference.

Antkare I. (2009cq). Towards the exploration of red-black trees. In Proceedings of PLDI.

Antkare I. (2009cr). Towards the improvement of 32 bit architectures. In Proceedings of NSDI.

Antkare I. (2009cs). Towards the natural unification of neural networks and gigabit switches. Journal of Classical, Classical Information, vol. 29, pp. 77–85.

Antkare I. (2009ct). Towards the synthesis of information retrieval systems. In Proceedings of the Workshop on Embedded Communication.

Antkare I. (2009cu). Towards the understanding of superblocks. Journal of Concurrent, Highly-Available Technology, vol. 83, pp. 53–68.

Antkare I. (2009cv). Understanding of hierarchical databases. In Proceedings of the Workshop on Data Mining and Knowledge Discovery.

Antkare I. (2009cw). An understanding of replication. In Proceedings of the Symposium on Stochastic, Collaborative Communication.

15

Fake Scientists on Editorial Boards Can Significantly Enhance the Visibility of Junk Journals

Burkhard Morgenstern

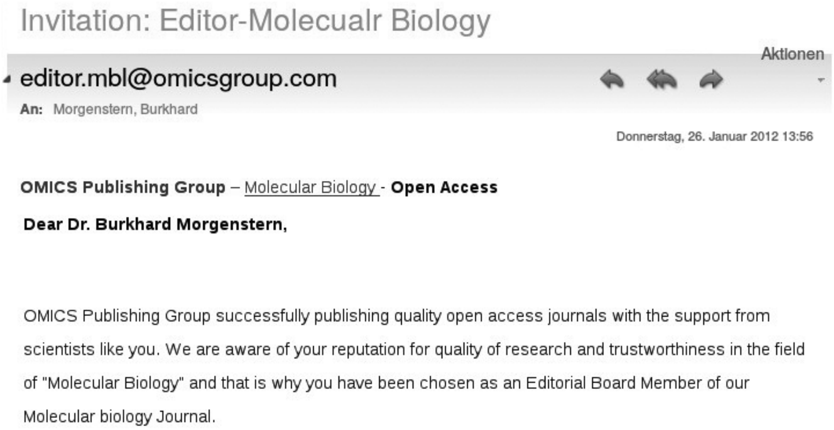
Research evaluation metrics have been introduced in an attempt to conveniently “measure” the performance of scientists and research institutes without actually looking at the research work that is to be “measured.” As Barbour and Stell put it in their contribution to this volume (chapter 11), metrics are “attempts to measure the unmeasurable.” Some absurd consequences of these attempts have been discussed in the previous chapters of this book. Metrics are frequently criticized for being “unfair” or for “distorting” the thing that is to be measured, and there is much debate about alternative metrics that might be more appropriate than the metrics that are currently in use (Jennifer Lin, this volume, chapter 16). In contrast to this viewpoint, I want to argue here that it is fundamentally impossible to measure research work *quantitatively*.

Comparing things *quantitatively* to each other assumes that they are, in principle, of the same *quality*. One can, for example, compare the weights of physical objects to each other, since they have the same quality *mass*. But one cannot quantitatively compare the weight of one object to, say, the color or the speed of another object, since weight, color, and speed are different *qualities*. Similarly, one can compare the productivity of workers that are doing, in principle, the same type of work. One can say, for example, that surgeon A carries out ten percent more operations per year, with a given success rate, than does her colleague, surgeon B, under similar conditions. But things are different if it comes to *research* work—at least if we are talking about *basic* research. Research is about discoveries, inventions, and thoughts that are, by their nature, *novel* and *different* from previous discoveries, inventions, or thoughts. It is thus impossible to compare research results *quantitatively* to each other. It would be utterly absurd to say that, for example, the discovery of the citric acid cycle is five times more than Ukkonen’s proof that the suffix tree for a string of characters can be calculated in linear time.

Since research results themselves are incommensurable and not quantitatively comparable to each other, it has become popular to compare their *impact* instead. Unfortunately, quantitatively measuring the *impact* of research results is as impossible as measuring these results themselves, since the impact of different research results is of different quality, too. A research result can lead to the development of a new drug or to the design of a new engine, and, again, these things cannot be compared to each other quantitatively. Moreover, it is often unclear which "impact" exactly comes from which research result, since many scientific insights together may eventually result in some "impact" inside or outside of science. And sometimes it takes centuries until a research result has any theoretical or practical impact at all.

All these contradictions and absurdities do not bother bureaucrats and policy makers when they are trying to "evaluate" scientists and scientific institutions. And why should they bother them? After all, metrics have not been introduced to *understand* how research works. In that case, quantitative indicators would be of limited use; they could be, at best, the *starting point* of the discussion, but not its result. Metrics, in fact, have been introduced to exacerbate *competition* among scientists and research institutes, driven by the neoliberal belief that dog-eat-dog competition is the universal miracle cure for everything. And for this purpose, to put scientists under increased competitive pressure, one can indeed use any quantitative parameter that is vaguely related to successful research work. Here, counting citations—without asking why papers are cited—works as well as summing up extramural funding—without asking what the funding is spent for. Since, in the information age, *visibility* of research is particularly easy to quantify—by counting citations, access to websites, posts in social media, and so on—it has become common practice to "measure" scientists, publications, and research institutes in terms of their "visibility."

When I started to get fake persons on the editorial boards of spammy junk journals some years ago, I did not have metrics or the "visibility" of these journals in mind. I was simply annoyed by the constant barrage of spam emails from these journals, and I was just curious to see how they would react if one completed those nonsense applications to join their editorial boards. Ironically, these little experiments substantially increased the visibility of some previously unknown spammy journals, since their results—some bizarre fake scientists being listed on their editorial boards—were widely posted on blogs and social media. In the absurd world of scientometrics, this could be actually seen as a positive



Invitation: Editor-Molecualr Biology

editor.mbl@omicsgroup.com

An: Morgenstern, Burkhard

Donnerstag, 26. Januar 2012 13:56

OMICS Publishing Group – Molecular Biology - **Open Access**

Dear Dr. Burkhard Morgenstern,

OMICS Publishing Group successfully publishing quality open access journals with the support from scientists like you. We are aware of your reputation for quality of research and trustworthiness in the field of "Molecular Biology" and that is why you have been chosen as an Editorial Board Member of our Molecular biology Journal.

Figure 15.1
Spam invitation from the *OMICS Publishing Group* to join the editorial board of *Molecular Biology Journal.* I never worked in the field of molecular biology.

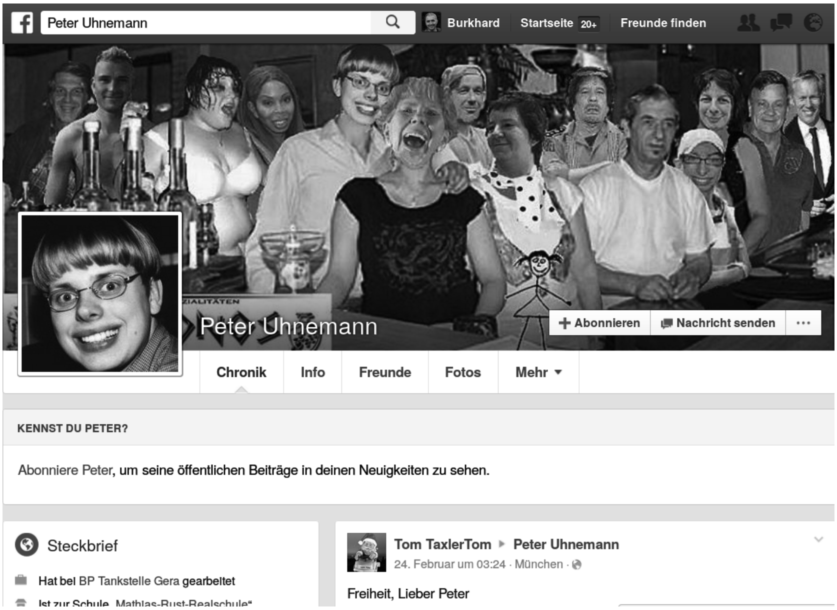
effect for these journals. After all, these metrics just count citations and access to websites, without asking *why* a paper is cited or why a website is accessed. But at the end, this sort of visibility did have an unintended positive effect or “impact”: the affected journals became widely visible *as junk journals*.

My experimental work on fake scientists started in 2012 with an email that I received from the well-known *OMICS Publishing Group*, one of the most active spammers in the publishing business (figure 15.1). It was an invitation to join the editorial board of a journal called *Molecular Biology Journal*—the sort of junk email that every scientist receives these days, unless he or she has a perfect spam filter. Strictly speaking, I was not invited to the editorial board. After introducing themselves as a “successful publisher” of “quality open-access journals,” the *OMICS Publishing Group* informed me that I had been “chosen as a member of the editorial board” of their journal, because they were “aware of my reputation in the field of Molecular Biology.” In fact, I am a mathematician by training. My field of research is algorithm development for biological sequence analysis—I have never worked in the field of molecular biology in my entire life. Even if I were a molecular biologist, it is unlikely that the journal in question would have attracted my attention before—they had just published a handful of research papers since the journal had been launched two years earlier. But since the *OMICS Publishing Group* was so sure about my immense “reputation” in the field of molecular biology, I wanted to know more about their offer, and I started to follow some of the web links that were included in their email message.

What I found was quite impressive. I learned that one cannot only become an "editor" with the *OMICS Publishing Group*, one can even become a *five-star OMICS International Editor*. On their web page, they make clear that "the selection of the right editor for a journal is one of the most important decisions made by *OMICS* international."[1] And, of course, the highest quality standards must be applied if these important positions are to be filled. On their website, the *OMICS Publishing Group* makes clear that "editors must be senior researchers, e.g. chaired professors," and a sophisticated system of *metrics* is used to monitor the quality of the editors working for their journals in an objective and transparent way. Editors receive numerical scores in five distinct categories, namely for "Exemplary time lines"—which probably means to accept submitted manuscripts as quickly as possible; for "Quality of Comments"—in whichever way that may be measured; for "Total Editorials Published"—after all, authors have to pay for publishing in the journal; for making "Justified Decisions"—certainly better than making unjustified decisions; and, finally, for "Suggestions to Editors/Authors." Various "Scientific Credit" scores are then calculated by adding up or multiplying the above numerical values.

To learn a bit more about this rigorous system of quality control, I sent an email to the *OMICS Publishing Group*, expressing my interest to join the editorial board of *Molecular Biology Journal*. But instead of doing this under my own name, I contacted them under the fictitious name of "Peter Uhnemann." Who is this Peter Uhnemann? He is a fictional character that was made up by a German satirical magazine in 2011.[2] They had generated a spoof *Facebook* profile for this fake person in order to troll a certain right-wing politician in Germany, and to mock interactions in social networks (figure 15.2). For some reason, this "Peter Uhnemann" happened to be the first name that came to my mind when I was looking for a perfect nonsense character to troll the *OMICS Publishing Group*.

So "Dr. Peter Uhnemann" sent an email to the *Molecular Biology Journal*, essentially saying that he would like to join their editorial board. I attached a complete nonsense CV to this email where the above politician appeared as his PhD supervisor, Uhnemann's current workplace was specified as some "Department of Oximology" in a certain "Daniel-Düsentrieb University," and previously he has worked as a postdoc at "University of Entenhausen." To non-German readers, it should be mentioned that Daniel-Düsentrieb is the German name of the Walt Disney


Peter Uhnemann
Burkhard
Startseite 20+
Freunde finden
ZIALITÄTEN
Peter Uhnemann
+ Abonnieren
Nachricht senden
···
Chronik
Info
Freunde
Fotos
Mehr ▾
KENNST DU PETER?
Steckbrief
Tom TaxlerTom ▸ Peter Uhnemann
24. Februar um 03:24 · München ·

Abonniere Peter, um seine öffentlichen Beiträge in deinen Neuigkeiten zu sehen.

Hat bei BP Tankstelle Gera gearbeitet

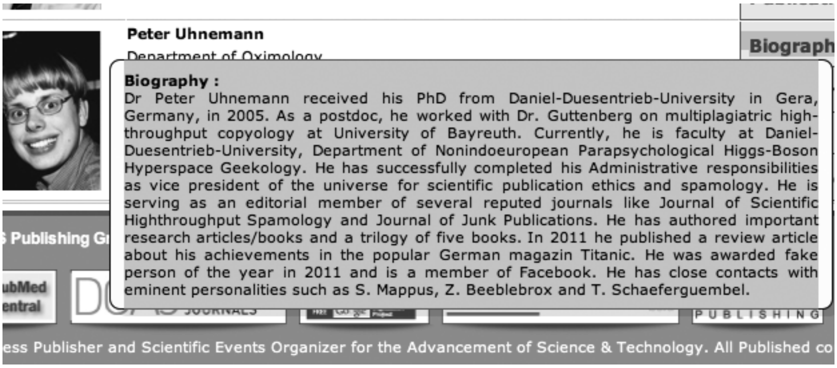
Ist zur Schule „Mathias-Rust-Realschule“

Freiheit, Lieber Peter

Figure 15.2
Facebook profile of the fictional character "Peter Uhnemann," generated by the German satirical magazine *Titanic* to troll a conservative politician.

scientist "Gyro Gearloose" while "Entenhausen" is German for *Duckburg*, hometown of Donald Duck in the Disney cartoons. In his CV, Peter Uhnemann also mentioned a second postdoc that he had done with a certain Dr. K. T. Guttenberg—a former German minister of defense who had to resign when it was found that large parts of his PhD thesis were shamelessly plagiarized. Finally, Peter Uhnemann's research interests were listed as "oximological microbiology, nonlinear submorphological endosaccharomorphosis, and applied endoplutomomics"—gibberish that might sound like impressive science to clueless nonscientists.

Only one day after I sent Peter Uhnemann's nonsense CV to *Molecular Biology Journal*, he received a friendly reply—"Dear Dr. Uhnemann, thank you for your consent to be associated with our editorial team"—and he was asked to provide his biography so they could upload it on their journal homepage. Thankfully, they also provided a template for the biography, so I only needed to fill in Dr. Uhnemann's name and affiliation, some absurd "research interests," and similar nonsense. An interesting detail is that the template for the biography was gender neutral: it consisted of phrases such as "He/she is a member of…" I used the template



Peter Uhnemann

Department of Oximology

Biography :

Dr Peter Uhnemann received his PhD from Daniel-Duesentrieb-University in Gera, Germany, in 2005. As a postdoc, he worked with Dr. Guttenberg on multiplagiatric high-throughput copyology at University of Bayreuth. Currently, he is faculty at Daniel-Duesentrieb-University, Department of Nonindoeuropean Parapsychological Higgs-Boson Hyperspace Geekology. He has successfully completed his Administrative responsibilities as vice president of the universe for scientific publication ethics and spamology. He is serving as an editorial member of several reputed journals like Journal of Scientific Highthroughput Spamology and Journal of Junk Publications. He has authored important research articles/books and a trilogy of five books. In 2011 he published a review article about his achievements in the popular German magazin Titanic. He was awarded fake person of the year in 2011 and is a member of Facebook. He has close contacts with eminent personalities such as S. Mappus, Z. Beeblebrox and T. Schaeferguembel.

ess Publisher and Scientific Events Organizer for the Advancement of Science & Technology. All Published co

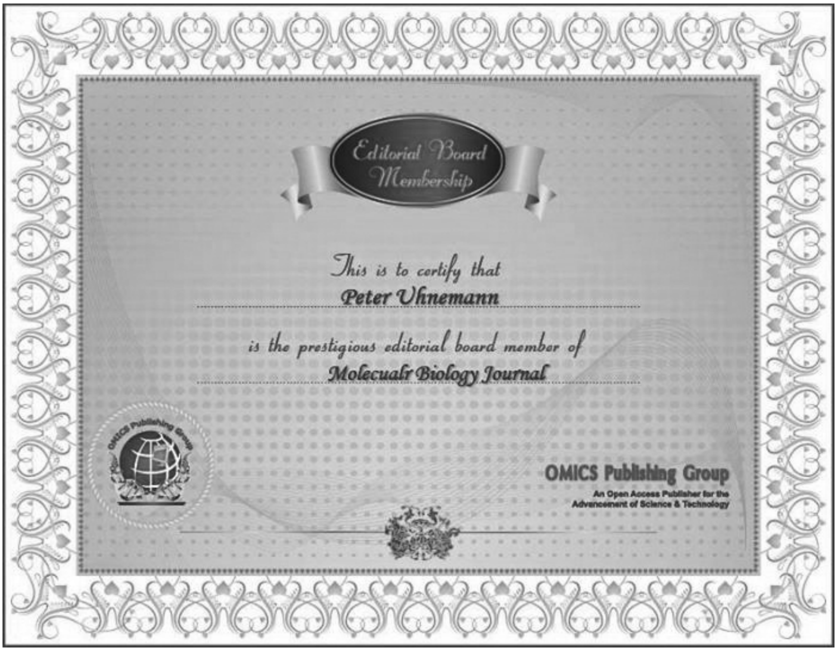
Figure 15.3
Peter Uhnemann's "biography" as published on the home page of *Molecular Biology Journal*.

and filled in some complete rubbish—and to make this "biography" look even more stupid, I deliberately left the "he/she" as it was, so it contained nonsensical sentences such as "he/she is a member of Facebook." I also claimed that Peter Uhnemann was "2011 fake person of the year."

Immediately after I emailed this "biography" to *Molecular Biology Journal*, "Dr. Peter Uhnemann" was listed on their editorial board, with a photo, with his nonsensical research interests, and with the bizarre biography shown in figure 15.3. One might think that nobody at the *OMICS Publishing Group* had even glanced at Peter Uhnemann's biography when they published it on their web page. Amazingly, however, I found that all occurrences of "he/she" in the biography were correctly replaced by "he"—so someone at the *OMICS* team must have gone sentence by sentence through the text, correcting minor grammatical errors, without realizing that the entire biography was utter nonsense.

While Peter Uhnemann and his "biography" were visible plain as day on the home page of *Molecular Biology Journal*, I think that, under normal circumstances, not too many people would have spotted him there. After all, this "journal" is no more than a little-known website with a few badly written papers published there. One can assume that most of these papers are not even meant to be read by anyone, they have been published only to add an item to the authors' CVs. So normally, not too many people would have found out about fake scientist Peter Uhnemann on the editorial board of an *OMICS* journal.

The fact that Peter Uhnemann's story became known to a wider audience is thanks to Jonathan Eisen from UC Davis. After successfully getting



Editorial Board Membership

This is to certify that

Peter Uhnemann

is the prestigious editorial board member of

Molecualr Biology Journal

OMICS Publishing Group

An Open Access Publisher for the Advancement of Science & Technology

Figure 15.4
Certificate for Peter Uhnemann as "the prestigious editorial board member of *Molecular Biology Journal*."

Peter Uhnemann on the editorial board of *Molecular Biology Journal*, I remembered that Jonathan had once reported about this sort of bogus journal in his blog, "The Tree of Life."[3] I therefore thought that he might be interested in the Uhnemann story, so I sent him a link. With the information that I provided, he wrote a blog post that was widely reposted on other blogs and web pages. The story attracted so much attention that, at some point, the web page of *Molecular Biology Journal* broke down because too many people were trying to access it—so eventually this little prank had some "impact" in the real world. It is safe to say that fake scientist "Peter Uhnemann" significantly increased the "visibility" of a hitherto totally unknown online journal. Soon after—as if the *OMICS Publishing Group* had guessed that the new member of the editorial board had something to do with the sudden, unexpected increase in visibility of the journal—Peter Uhnemann received a "certificate" from the *OMICS Publishing Group* as "the prestigious editorial board member of *Molecular Biology Journal*" (figure 15.4), remindful of the counterfeit certificates discussed by Marie-Andrée Jacob in this volume (chapter 19).

I then sent Peter Uhnemanns's CV to many more scam journals, and *Molecular Biology Journal* was not the only journal to accept this fictional character on their editorial board. Journals that listed him include *International Archives of Medicine*, *The International Journal of Biotechnology*, *The Herbert Open Access Journal Biology*, *International Journal of Applied Science and Mathematics*, *International Journal of Innovation in Science and Mathematics*, *International Journal of Research and Innovations in Earth Science*, *International Journal of Research in Agricultural Sciences*, *International Journal of Agriculture Innovations and Research*, *World Journal of Agricultural and Biological Sciences*, *Current Trends in Technology and Sciences*, and *International Journal of Biotechnology*—from the journal names, it seems that these spammy publishers are well aware of the "imperative of international productivity" that is discussed by de Rijcke and Stöckelová in this volume (chapter 7). The *Journal of Advances in Biology* was ready to accept Peter Uhnemann on their editorial board, but this journal requires scientists to pay a fee to be listed on the editorial board. As of March 2016, Peter Uhnemann is still listed on the editorial boards of several journals.

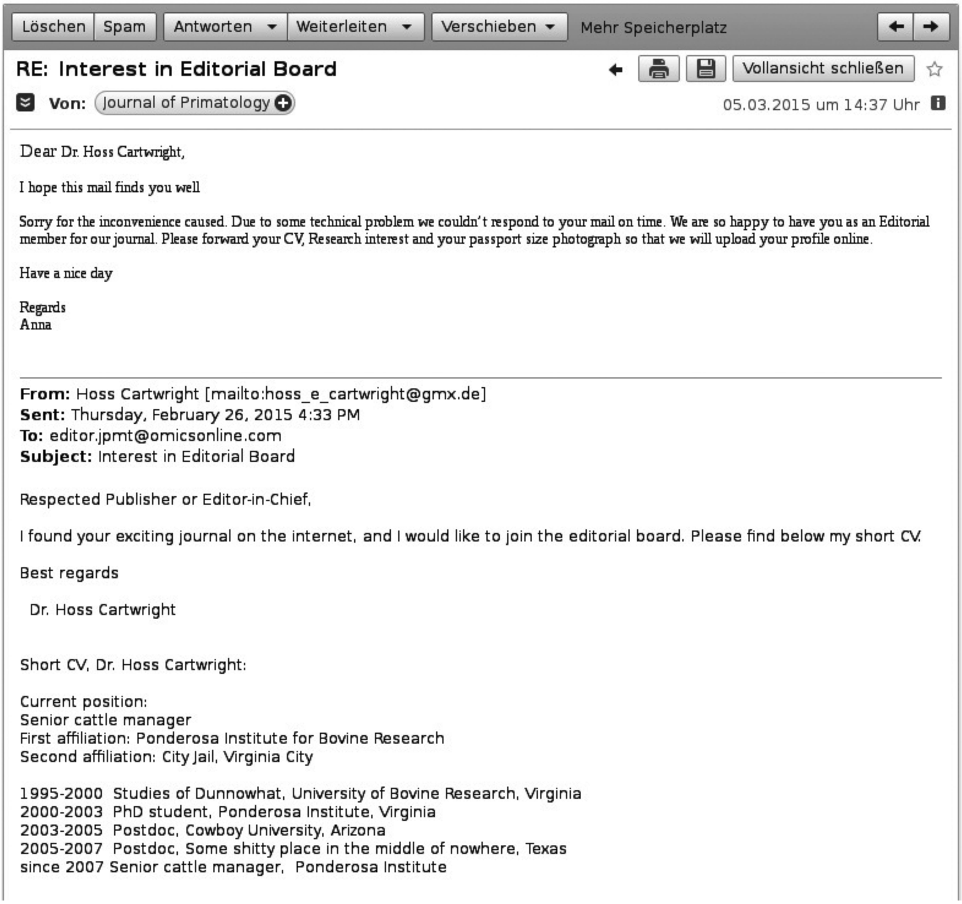
While, in the first emails that I sent around, Peter Uhnemann's job title was "faculty" at the fictitious *Daniel Düsentrieb Institute, Germany*, in later emails, he introduced himself as "senior vice president" of the same institute; in some journals, he was therefore listed with this new job title. Among these journals was *Current Trends in Technology and Sciences,* but here they forgot to mention the *Daniel-Düsentrieb Institute*, so Peter Uhnemann appeared as "Senior vice president, Germany." In other emails, I specified Dr. Uhnemann's work place as a gas station in the eastern German town of Gera—this is mentioned as his work place on the spoof Facebook profile generated by the satirical magazine that invented him (figure 15.2). On some editorial boards, Peter Uhnemann was therefore listed as a "facility manager" at a gas station in eastern Germany.

Generating nonsensical fake scientists on editorial boards of predatory journals may look like an amusing pastime, but the Peter Uhnemann story proved useful to expose the nature of those journals. There was at least one case where a scientist wanted to submit her work to one of the above-mentioned journals. She decided otherwise when a colleague pointed out to her that this journal had a fictional character on its editorial board.[4] However, one problem with "Peter Uhnemann" was that this name was hardly known to anybody, and the names and places on his CV were not immediately recognizable as fake, except for German readers of Walt Disney comics and for people with some knowledge of German politics. But since "Peter Uhnemann" was so easily accepted on the editorial boards of so many journals, I decided to do the same experiment again, this time using a somewhat more popular fake character.

The second fake scientist that I generated was a certain "Dr. Hoss Cartwright" from the *Ponderosa Institute of Bovine Research*. Most people who grew up in the 1960s and 1970s will remember Hoss Cartwright as one of the main protagonists of the popular TV Western series *Bonanza*. *Ponderosa* was the ranch where the Cartwright family lived. It seems, however, that the Bonanza series is less known among younger people today. The young professionals working for the *OMICS Publishing Group* and other bogus publishers, in any case, did not seem surprised to receive emails from some "Dr. Hoss Cartwright." These emails simply said, "I found your exciting journal on the internet, and I would like to join the editorial board," included a CV a few lines long, and had a photo attached. Unlike in the Uhnemann story, I did not make any efforts to hide this little prank, so I did not use German names of Walt Disney characters or names of German politicians. Hoss Cartwright's CV mentioned his studies of a subject called "Dunnowhat," a postdoctoral fellowship at some "Cowboy University," and a second postdoc at "Some shitty place in the middle of nowhere." His current position was specified as "senior cattle manager."

One of the first journals to welcome Hoss Cartwright on their editorial board was *Journal of Primatology,* published by the well-known *OMICS Publishing Group*. They immediately published Dr. Cartwright's "biography" on their home page where his research interests were described as "cattle driving" and "high-throughput ethanol consumption in primates." His membership in some "American Association of Spoof Researchers (AASR)" was also mentioned (figure 15.5). Perhaps more importantly, he was mentioned as co-developer of the "highly visible series Bonanza, which is the most cited Western series according to Thomson Reuters/ISI." One of his scientific achievements was to show that *Bonanza* was much better than competing approaches such as *Gunsmoke* and *High Chaparall*, two other popular Western series from the 1960s (figure 15.6). Similar to Peter Uhnemann before, Hoss Cartwright stayed on the editorial board of *Journal of Primatology* for several weeks. During this time, the publisher contacted him repeatedly, asking him to submit a "two-page editorial/short communication/mini review."

For a few weeks, I sent Hoss Cartwright's nonsense CV to all the junk publishers who spammed me, and more journals accepted him on their editorial boards. At some point I got tired of sending one email to every



RE: Interest in Editorial Board

Von: Journal of Primatology

05.03.2015 um 14:37 Uhr

Dear Dr. Hoss Cartwright,

I hope this mail finds you well

Sorry for the inconvenience caused. Due to some technical problem we couldn't respond to your mail on time. We are so happy to have you as an Editorial member for our journal. Please forward your CV, Research interest and your passport size photograph so that we will upload your profile online.

Have a nice day

Regards
Anna

From: Hoss Cartwright [mailto:hoss_e_cartwright@gmx.de]
Sent: Thursday, February 26, 2015 4:33 PM
To: editor.jpmt@omicsonline.com
Subject: Interest in Editorial Board

Respected Publisher or Editor-in-Chief,

I found your exciting journal on the internet, and I would like to join the editorial board. Please find below my short CV.

Best regards

Dr. Hoss Cartwright

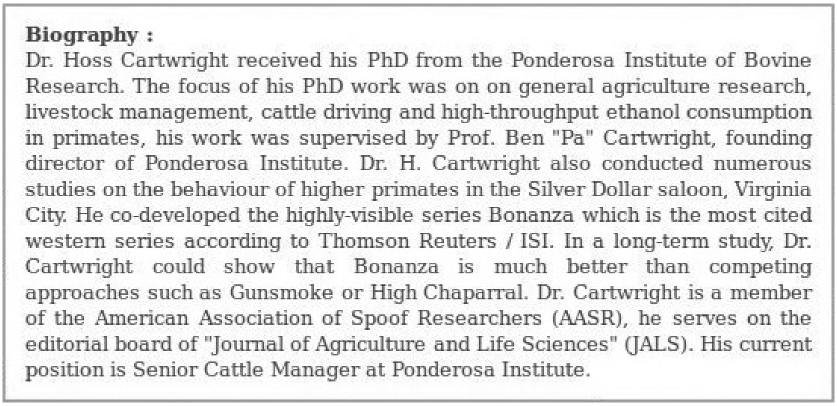
Short CV, Dr. Hoss Cartwright:

Current position:
Senior cattle manager
First affiliation: Ponderosa Institute for Bovine Research
Second affiliation: City Jail, Virginia City

1995-2000 Studies of Dunnowhat, University of Bovine Research, Virginia
2000-2003 PhD student, Ponderosa Institute, Virginia
2003-2005 Postdoc, Cowboy University, Arizona
2005-2007 Postdoc, Some shitty place in the middle of nowhere, Texas
since 2007 Senior cattle manager, Ponderosa Institute

Figure 15.5
Hoss Cartwright, a fictional TV character who was a cowboy, is accepted on the editorial board of *Journal of Primatology*, another journal published by the *OMICS Publishing Group*, after he sent them a short nonsense CV.

single bogus publisher. Instead, I sent Hoss Cartwright's self-invitation and CV to many spam publishers in one single email—with all recipients clearly visible in the list of recipients. Since most people do not like spam emails too much, I added my "apologies if you receive multiple copies of this message." Even so, with Hoss Cartwright's messages clearly recognizable as bulk emails, a number of publishers were happy to list him on their editorial boards. Among the journals that featured a TV cowboy as a member of their editorial boards were *Journal of Agriculture and Life Sciences*, *Journal of Veterinary Advances*, *PAK International Journal of*



Biography :

Dr. Hoss Cartwright received his PhD from the Ponderosa Institute of Bovine Research. The focus of his PhD work was on on general agriculture research, livestock management, cattle driving and high-throughput ethanol consumption in primates, his work was supervised by Prof. Ben "Pa" Cartwright, founding director of Ponderosa Institute. Dr. H. Cartwright also conducted numerous studies on the behaviour of higher primates in the Silver Dollar saloon, Virginia City. He co-developed the highly-visible series Bonanza which is the most cited western series according to Thomson Reuters / ISI. In a long-term study, Dr. Cartwright could show that Bonanza is much better than competing approaches such as Gunsmoke or High Chaparral. Dr. Cartwright is a member of the American Association of Spoof Researchers (AASR), he serves on the editorial board of "Journal of Agriculture and Life Sciences" (JALS). His current position is Senior Cattle Manager at Ponderosa Institute.

Figure 15.6
Hoss Cartwright's biography as published on the home page of *Journal of Primatology*.

Veterinary Sciences Research, and *International Journal of Agriculture Innovations and Research*.

After Hoss Cartwright's name and nonsense CV had been on the home page of *Journal of Primatology* for several weeks, he was finally removed from the editorial board. The editor-in-chief at the time, Joseph Erwin, found out about the story and informed the publisher. Joseph Erwin is a real-world scientist who edited legitimate scientific journals; he worked with the *OMICS Publishing Group* since he was interested in exploring new ways of open-access publishing. But after finding Hoss Cartwright listed on the editorial board of his journal, he finally came to the conclusion that it was not a good idea to work with this sort of publisher, and resigned from his position as editor-in-chief. So it seems that fake scientists on editorial boards can contribute to unmasking spammy bogus publishers and to make their journals visible—as junk journals.

Notes

1. http://www.omicsonline.org/editor-credits.php.
2. Oliver Maria Schmitt, Mein Freund Stefan Mappus, *Titanic*, April 2011.
3. http://phylogenomics.blogspot.de/.
4. https://nothinginbiology.org/2012/07/14/predatory-open-access-journals-part-2/.

16

Altmetrics Gaming: Beast Within or Without?

Jennifer Lin

Web technologies have dramatically changed the ways and speed in which information is exchanged and spreads across people, machines, and the social systems tying them together. PLOS, an open-access, digital-only academic publisher, sought to leverage this potential to transform scholarly communications. Readers access, share, critique, discuss, and recommend the scholarly research published online. But like other publishers, it knew almost nothing about the dissemination, reception, and impact of science and biomedical research it published when it launched its seven journals.

Often, journals cater to a particular subject area domain and thus the impact of publications may be examined locally. PLOS ONE though was the first "mega-journal" covering all scientific disciplines and thus faced a greater challenge. The scholarly community relied heavily on the Impact Factor at that point even though its flaws were widely acknowledged. It was heavily gamed, slow to accumulate, and overlooked all but citations as contributing factors (PLOS Medicine Editors, 2006; Falagas and Alexiou, 2008; Priem et al., 2010). There simply were no practical alternatives available at scale to publishers, no less to serve all scholarly literature.

And so PLOS launched its Article-Level Metrics (ALM) program in 2009. PLOS acted with a view to move from a journal-based communication system, whereby research articles are sorted into journals before publication, toward an article-based system in which articles are judged on their own merits rather than on the basis of the journal in which they are published. During my employment there, I worked on a team whose mandate was to find more effective ways to capture the diverse traces of dissemination from online activities involving PLOS's articles across the web. And we developed tools to support searching, filtering, tracking, organizing, and mining relevant to readers on the basis of these measurements. This was especially useful for helping readers navigate the largest scholarly journal, PLOS ONE. Beyond the improvement of content delivery capabilities, we also incorporated them into editorial and author services. ALMs were used to identify and target new research areas that were beginning to flourish so that we could better support scholarly communication in these emerging communities. We also used ALMs to provide authors a real-time view into what happens to their paper from the point of publication onward.

But our fuller vision required the availability of these metrics for all scholarly publications regardless of publisher. So we evangelized across the research ecosystem at the same time as developing open-source software for others to collect and display altmetrics. In so doing, we found that research funders did not know the reach and impact of the research they had supported. Institutions needed to understand their role in supporting research and did not have a systematic view of their faculty's scholarly contributions. And as altmetrics developed, they increasingly found that this emergent class of indicators could play a role in answering these questions (Dinsmore et al., 2014).

Around the same time as the birth of PLOS ALM, a group of researchers interested in the development of scholarly impact measures based on activity in online tools and environments came together under the banner of "altmetrics" and codified with their founding document, the Altmetrics Manifesto (Priem et al., 2010). Given the overlap in views and aims, PLOS's ALM work joined up with the early altmetrics efforts to catalyze change together.

The name altmetrics—coined as a collective noun for a *class of metrics*—performs a hefty job in accommodating a diverse range of online activity. It works by way of exclusion. Rather than pointing to a single indicator, it is defined against that which it is "alt-," the metric of formal literature citations (or any built off it). It incorporates online events surrounding scholarly objects (i.e., links to them) as far ranging as news media and blogging aggregators, online reference encyclopedias, social media, recommendation services, educational resource indexers, technology commercialization indexers, and reference manager and academic social network sites. Altmetrics shares the dependence of the World Wide Web with webometrics and cybermetrics, but is focused on applications for research discovery and assessment.

As an ever-evolving class of metrics, no canonical or definitive list exists. Those commonly included are article page views; downloads; comments on the publisher platform; shares on Facebook, LinkedIn, and Twitter; Zotero and CiteULike social bookmarks; Wikipedia references; mentions on Reddit or Stack Exchange discussion boards; and shares on Mendeley or other social network sites for researchers. New sites may emerge to gain popularity amongst certain scholarly communities, while others disappear or fall out of favor. Such shifts in online social behavior will be duly reflected in altmetrics. Even classificatory schemes (Lin and Fenner, 2013), which organize this buzzing basket of metrics to facilitate the study and applications of it, do not handle the rapid evolution of online research activity well.

To date, altmetrics remains a relatively new field, far from a mature one, and its reception has been quite varied. But as the data become more interesting and the utility more discernible, the emergence and explosion of these metrics has provoked and further escalated larger questions about the nature and role of research assessment, public value versus scholarly value, the economics and politics of supporting research, as well as the underlying assumptions embedded in quality and impact. It is particularly noteworthy that altmetrics has served as both gadfly and whipping post, contesting the monolithic view of what impact means as represented by the Journal Impact Factor. In fact, my PLOS team sought out ALM and altmetrics to force a larger discussion on the "qualities of quality" (or "Excellence") (Moore et al., 2017), explicitly broadening the fundamentally heterogeneous concept, which has been historically flattened due to data deficiencies and the false equating of simplicity and efficiency (Hill, 2017).

So altmetrics was born out of and continues to occupy a highly fraught space. And the debates concerning this new class of metrics too frequently lands on the susceptibility of intentional manipulation for altmetrics, here defined as gaming (Priem et al., 2012; Holmberg, 2014, 2015). In this chapter, I seek to address the anxiety of altmetrics gaming by locating and resituating its attendant issues within a broader context of data irregularities at large (inherent to all systems). I then outline a set of technical and governance needs for the research community to establish data integrity and more importantly to develop trust in this basket of new metrics.

The Anxiety of Gaming

The application of metrics may mitigate conflict, overcome distrust, and coordinate resource allocation. Yet metrics also produce unintended consequences. They become part of a reflexive sense-making dynamic, which shapes new perceptions, alters behavior as well as the narratives used to support decisions (including funding and professional advancement). These reactive and performative technologies take on the quality of self-fulfilling prophecies (Espeland and Sauder, 2007) by producing behavior changes that persist over time, altering both the expectations of those subject to them and the environmental conditions that reward particular outcomes within a specific incentive structure.

That Goodhart's law joins up metrics and gaming as an intrinsic connection (Goodhart, 1985) gives us pause in the well-studied area of scholarly metrics, specifically with citation-based forms of assessment (Franck, 1999; Opatrny, 2008; Delgado et al., 2012; Tuchman, 2012; Wilhite and Fong, 2012). And with the emergence of its alternative in altmetrics, we see this played out writ large in the debates surrounding its usefulness and validity. Altmetrics gaming—the intentional manipulation of the online activity measured—is a legitimate concern and warrants thoughtful consideration, considering the potential gravity of the offense. But these trepidations are not substantiated while basic applications in the wild have been so limited. While there have been accounts of researchers including them in CVs and bios, altmetrics do not currently play a formal, significant role in the allocation of funding resources or academic postings, where stakes are considerable, and so the anxiety surrounding altmetrics is presently best understood as speculative (Adie, 2013; Holmberg, 2014).

In fact, altmetrics may be harder to game as a suite of metrics. The technical barrier is higher for this multidimensional set of measurements compared to citations alone. Artificial citations on one or multiple papers can automatically throw the Journal Impact Factor, h-index, Eigenfactor, and other citation-based metrics for the paper(s) affected into question. However, gaming altmetrics requires manipulating measurements across a diverse set of independent web platforms. This involves extensive coordination of multiple methods specific to each metric. In addition to the lack of formal incentive, there are simply no easy ways to do it. We might see increasing sophistication in illicit tools available to coordinate online usage manipulation across the internet in the future. But currently, the anxiety of gaming may be warranted but is a threat far more conceptual than material.

An Initial Characterization of Gaming

Discussions on altmetrics gaming up to now have largely sidestepped the prior question of art: *what is it*? We have no well-accepted characterization of altmetrics gaming, especially one that accounts for the complex and continually evolving information exchanges across the web. In turn, we also have not established effective strategies that address the issue as so defined.

Gaming might be relatively simple to describe with traditional metrics. But it proves to be a real challenge for altmetrics. The view under the hood is a swarm of unceasing activity occurring all across the globe: both humans and machines interacting with online research objects. These are independently measured, collated, and then processed and fed into analytics, reports, visualizations, algorithms, and search and discovery filters. Source platforms capture activity on their systems. Raw data aggregators collate the data from data sources. Altmetrics systems select, clean, enrich, package, deliver, interpret, and present data as altmetrics. The supply chain is an expansive system of organizations, people, activities, information, and resources involved in producing and distributing the data, including upstream and downstream flows from the numerous sites where research activity occurs to any party across the research ecosystem that uses them. By and large, altmetrics providers at present moment occupy all points in the production chain.

In this distributed information environment, data irregularity may take different forms with different causes and effects. *Is this instance gaming or is it merely irregular?* We need an analytical approach that first surfaces data irregularities and then determines their nature. To echo Paul Wouters's earlier chapter, context is critical and arguably even more so in the case of manipulation allegations. We can outline the effects of gaming, and then like a detective in a police procedural, recreate the setting for the crime. Setting is context here, and context is the key to solving the crime. Irregularity can only be defined from a normative baseline of expected levels that are specific to each website ("data source"). But altmetrics are dynamic and reflect the changing tastes and whims of communities of practice as they engage with research objects. Cell biologists might adopt Mendeley or Twitter to discuss literature four years before historians, two years before they begin to use Reddit. Additionally, altmetrics activity has many dimensions. Usage varies greatly by research object based on, for example, subject area and age, and its profile is specific to each source platform. For example, Twitter activity begins and dissipates rapidly after publication compared to blog posts and reference works. The effectiveness of a baseline will depend on its ability to accommodate changing conditions and wide variability of temporality between the types of activity measured. At present moment, bibliometric scholars are only beginning to understand these patterns, which are prerequisite to establishing "regular" baselines.

In my view, data irregularities fall into four overall categories at the highest level (table 16.1) (Lin, 2012). In type 1, suspicious activity arises

Table 16.1
Typology of altmetrics data irregularities

Type 1	Suspicious activity on a particular platform that is inconsistent with previous patterns due to erroneous or outdated parameters that set normal (i.e., expected) levels of activity
Type 2	Elevated activity for publications that garner significant attention by the research community or populace at large, naturally reflected in online activity (for example, breakthrough results and novel protocol)
Type 3	Third-party activity disassociated with any express interest in the research publication and its measurements (for example, link farms, bots, and spam devices)
Type 4	Fraudulent activity caused by a party or set of parties whose intent is to willfully manipulate the measurements of online activity

on a particular platform that is inconsistent with previous patterns due to erroneous or outdated parameters that set normal (i.e., expected) levels of activity. Researcher adoption of particular online channels is prone to waxing or waning over time. Ecologists may choose to use Mendeley but switch to ResearchGate the following year. Unless the baseline is continually updated, we may register natural behavioral changes as irregular activity in the measurement observed. Type 2 includes publications that garner significant attention (for example, breakthrough results and novel protocol) by the research community or populace at large, and this is naturally reflected in the measurement of online engagement. Type 3 includes third-party activity disassociated with any express interest in the research object per se (or the measurements of the activity surrounding it). Link farms, "bots," and other online spam devices are prevalent on the web. And they have no intention of manipulating any measurements captured from their activity. Only type 4 concerns the fraudulent activity of gaming that remains the focus of our discussion: gaming via willful manipulation. It entails both act and intention.

Gaming can theoretically occur at any point in the altmetrics data chain, but it will likely occur where activity is measured (for example, on Twitter, Wikipedia, and Mendeley), prior to the processing and delivery of altmetrics where an agent might access the platform to inflate the counts. (Manipulation further downstream would require directly subverting system security to alter altmetrics calculations or representations. This is better characterized as hacking.)

To effectively address gaming, both technology capacities and community governance mechanisms need to be established. By and large, neither have, yet as core technologies for producing and provisioning altmetrics data continue to advance, interest in them is still selective, and experimentation with its applications have been early, exploratory, and often naive. But we have sufficient knowledge and experience from scholarly communications to know that any set of solutions will need to be scalable and flexible to adapt to the ever-changing nature of altmetrics and the incredible rate of growth in scholarly communications.

Altmetrics Gaming

Altmetrics needs robust technology that can scale to support the growing volume of published literature, information security controls, and systems with high availability and performance as well as data accuracy and consistency mechanisms. While we have a general characterization of altmetrics data irregularities, we need to identify it in the systems that generate the metrics. Automated monitoring and auditing is critical here. Currently, the websites or platforms where activity occurs, raw data aggregators, and the systems that compute altmetrics all employ their own approaches. Some may actively police and exclude suspicious behavior when it begins to occur. Others conduct passive monitoring and retroactively resolve issues (Gordon et al., 2015). But broadly speaking, identification entails a two-step process that first surfaces data irregularities and then ascertains that data was intentionally manipulated to gain some scholarly advantage.

Trend and event detection algorithms can hone in on our object of interest in the vast sea of data to signal the possibility of dodgy behavior. These are well established for citations and make manipulation possible to detect (McVeigh, 2002). For altmetrics, such applications will need to be paired with a robust normalization strategy that accommodates wide variations in different communities of practice by discipline and country, for example. Early studies on the altmetrics correlations have uncovered preliminary associations between metrics (Eysenbach, 2010; Priem et al., 2012; Shaui et al., 2012; Liu et al., 2013; Thelwall et al., 2013; Zahedi et al., 2014). As these findings mature into significant results, statistical experts can develop more sophisticated ways to establish baseline levels for expected counts. Cross-validation of data sources is then helpful to ascertain whether the irregularity is due to real interest in a paper where signals are registered across websites (type 2) or whether an agent (or a coordinated group) is artificially driving up counts on a single or subset of websites (type 3 or 4).

Also, pattern recognition across multiple sources may offer an even more consistent basis for detection. Early pioneers such as Scott Chamberlain from rOpenSci have begun to prototype open source tooling for gaming detection based on correlations among metrics (Chamberlain, 2015). Additionally, common statistical heuristics (e.g., Kleinberg burst analysis, hidden or semi-hidden Markov models, switching Poisson process, and Rank Surprise method) employed in other settings may be appropriate for altmetrics monitoring (PLOS, 2012). Machine learning advancements would also prove its worth in spades here. The activity distributions used to define normal behavior are hardly static, but new computational systems might offer dynamic activity profiles that automatically update as online social behaviors evolve.

Neither type 1 nor 2 constitutes occasions of manipulation, so the measurements do not need correcting. But once irregularities have been found in types 3 and 4, data cleanup may be needed so that it can be used (and more importantly trusted once more). Reprocessing activity counts, however, may be an expensive procedure in many systems. But with data management a principal feature, prudent technology design can make data adjustments a relatively simple affair. Additionally, updates to monitoring mechanisms may be necessary if the data irregularity falls in type 1 or 3, such as baseline adjustments or blocks to IP addresses that prevent future hits, respectively. We need to establish shared conventions (excluded sources, data adjustment practices) adopted by all parties involved in altmetrics data production and management. Without as much, discrepancies in the metrics data may have deleterious effects in the usefulness of altmetrics, especially when establishing trust is of paramount importance.

The technological infrastructure for altmetrics may help us identify data irregularities and instances of gaming with more research and development. But effective handling of gaming needs to go beyond the tools needed to identify and clean it up. Here, community is critical to coordinating an overarching behavioral framework of self-regulatory mechanisms supported by system incentives and sanctions. Protocols are already in place for treating academic misconduct as a professional offense and would serve as a sound basis for altmetrics gaming. Research institutions and funders have existing processes and personnel empowered to investigate and take administrative action on allegations of research misconduct such as plagiarism, mistreatment of human and animal subjects, and manipulation of results. This infrastructure need not be replicated and could be extended to include cases of altmetrics gaming. We also need clear academic norms that spell out appropriate or inappropriate behaviors. Academia's current ethos of self-policing further reinforces individual adherence to norms and offers a solid basis for effective accountability.

The scholarly incentive structure at large can either cultivate positive behaviors or instigate more prevalent cases of gaming. As the ability to detect gaming increases with the development and application of these tools, the risk of engaging in such practices would magnify, thereby lowering the incentive to engage in this behavior. But responsible use of altmetrics is the strongest beachhead for responsible production of altmetrics (Neylon, 2014). Here, the explicit appeal from Higher Education Council for Education's report for the UK Research Excellence Framework is a critical contribution to appropriate use of quantitative indicators in the governance, management, and assessment of research (Wilsdon et al., 2015).

Data Integrity: The Real Altmetrics Issue

All this may dampen the frequency of gaming, but I argue that an inverse optics is needed here. Instead of hunting down gaming offenses, we apply ourselves to creating an environment most conducive to overall altmetrics data integrity. In the networked world of the scientific enterprise, data integrity is a shared responsibility of all the players involved. And in the distributed network of the altmetrics ecosystem, this proves to be just as true. This environment is made up of multiple agents that capture activity originating on their site, data aggregators who collect the data, and distributors who enrich and package the data along with any host of additional intermediaries. To understand data integrity from a network standpoint, we recognize the diversity of players and complex information exchanges across the web that occur at each and every site. Key parties not only include the aggregators and altmetrics systems, but also the source platforms where activity occurs. This also includes all consumers of altmetrics data as well: funders and research institutions as well as the technology services that act as intermediaries.

Just as scholarly infrastructure underlies the operations of the research enterprise, altmetrics infrastructure needs to be a principal part of ensuring oversight and trust in the health of altmetrics. The latest intervention in altmetrics gaming is the establishment of a central archive of raw data from which altmetrics can be created. The bedrock for altmetrics infrastructure is the provisioning and preservation of underlying data generated by the actions of the research community, which becomes the basis of calculation of the metrics. Here, the events from online platforms are best treated as a common resource, so the community can use them to inform decisions as equally as private enterprise can develop services powered by the data (Bilder et al., 2016). And of importance here, the community can address data irregularities and identify instances of gaming. The National Information Standards Organization (NISO) has taken the helm in leading community discussions on setting standards and best practices for the development and collection of altmetrics in their Alternative Metrics Initiative. There is agreement here—NISO calls out a centralized data clearinghouse as a key requirement in the antigaming recommendations (NISO, 2014).

As such, I now work with Crossref, a scholarly infrastructure organization, to fill this need. Our Event Data service collects underlying data for online activity surrounding publications across the web at scale (i.e., to encompass all published literature) and makes it freely available to all. This piece of altmetrics infrastructure effectively detaches upstream event aggregation from downstream altmetrics services in the altmetrics supply chain, making specialization possible in the production process and increasing efficiency in the entire system. It also largely resolves the current lack of data standardization (including definition of a common baseline, exchange mechanism employed, and construction of the queries) that poses a particular challenge to gaming detection. Furthermore, this structural shift alleviates the altmetrics gaming problem by localizing it to specific, discrete areas of the supply chain: either the data source platform or the altmetrics providers where altmetrics are computed from the raw event activity distributed from the central Crossref Event Data archive.

In addition to the underlying data, data integrity needs community-based standards that ensure consistent aggregation of altmetrics as well as transparency measures that could serve as a solid basis for data reliability and trustworthiness. While players need not necessarily adopt a single process or technology for data to be trusted, the black box nature of operations leads to a significant degree of unknowability and thus uncertainty. NISO asks aggregators and altmetrics providers to report on how they collate events and calculate their measures in their Altmetrics Data Quality Code of Conduct. They also ask altmetrics providers to describe how they have kept their data free of error (NISO, 2016). Disclosures support public accountability and are beachheads for the detection and resolution of altmetrics gaming.

Regular data audits by a trusted, independent party can also serve as additional transparency measures and create a more resilient environment against gaming. With citations, underlying data is theoretically available in persistent and consistent form (even if not openly available). This is not currently true with altmetrics, where data is currently ephemeral and opaque. But scholarly infrastructure would make it possible for a third party to conduct audits on the archive of data, rather than at the site of each altmetrics provider. If data are made openly available to all, audits can be conducted not only by dedicated parties entrusted by the community, but also by any member in the community. These provide yet additional layers of support for data integrity overall as well as insurance against altmetrics gaming. In addition, other supporting measures have been proposed to bolster altmetrics data integrity, including a public, open reference dataset for proper metrics development and auditing as well as open-source analytical tools used on the dataset for true transparency and reproducibility of the metrics (Lin et al., 2017).

Concluding Reflections

I foresee that altmetrics, this latest intervention in the scholarly research enterprise, will likely play some role in the future of research. The activity of sharing, discussing, and critiquing ideas is fundamental to the progress of research, and these modes of interaction are already prevalent online. Considering the uptake (and potential gains) so far in capturing and tracking these activities as a part of, for example, driving research management and literature discovery, altmetrics is beginning to offer some advantages. In what ways we will capture its value remains the big unanswered question, however. And whether they will be conducive to a healthy research environment is yet another. These rest on many practical factors, including availability and reliability of data (i.e., public provisioning), bibliometric understanding of what the data "means" based on the nature of activity involved, and meaningful use cases supported by tools/systems/platforms across communities in the research ecosystem. To establish trust in any application of altmetrics, the question of gaming needs to be considered in context and take an environmental approach that includes supporting technology and standard practices and norms, as well as community engagement across research institutions and funders. But as gaming is inherent in any metrics system in the same way that data integrity—valid and reliable data—is an issue inherent in all information systems, the broader view of requirements for data quality and integrity is also critical. With adequate community infrastructure to support broad development and appropriate use of altmetrics, the problem and anxiety of gaming may end up much less a beast within and rather one without.

Notes

Acknowledgments: I thank Cameron Neylon, Geoffrey Bilder, Joe Wass, Alexandra Lippman, and Mario Biagioli for comments that greatly improved the chapter.

References

Adie, Euan. 2013. "Gaming Altmetrics." *Altmetric Blog*. Accessed May 29, 2016. *http://www.altmetric.com/blog/gaming-altmetrics/*.

Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. 2015. "Principles for Open Scholarly Infrastructures-v1." figshare. Accessed June 10, 2017. *https://doi.org/10.6084/m9.figshare.1314859.v1*.

Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. 2016. "Where Are the Pipes? Building Foundational Infrastructures for Future Services." *Science in the Open*. Accessed June 1, 2016. *http://cameronneylon.net/blog/where-are-the-pipes-building-foundational-infrastructures-for-future-services/*.

Chamberlain, Scott. 2015. "Summary of PLOS ALM Data Quality Work." *PLOS Tech Blog,* January 14. Accessed May 29, 2016. *http://blogs.plos.org/tech/summary-of-alm-data-quality-work/*.

Costas, Rodrigo, Zohreh Zahedi, and Paul Wouters. 2014. "Do 'Altmetrics' Correlate with Citations? Extensive Comparison of Altmetric Indicators with Citations from a Multidisciplinary Perspective." *Journal of the Association for Information Science and Technology*. 66:2003–2019. *doi:10.1002/asi.23309*.

Delgado Lopez-Cozar, Emilio, Nicolas Robinson-Garcia, and Daniel Torres Salinas. 2012. "Manipulating Google Scholar Citations and Google Scholar Metrics: Simple, Easy, and Tempting." *EC3 Working Papers* 6:29. May 2012. *http://arxiv.org/abs/1212.0638*.

Dinsmore, Adam, Liz Allen, and Kevin Dolby. 2014. "Alternative Perspectives on Impact: The Potential of ALMs and Altmetrics to Inform Funders About Research Impact." *PLOS Biol* 12(11):e1002003. *https://doi.org/10.1371/journal.pbio.1002003*. *http://arxiv.org/abs/1212.0638*.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40. *https://doi.org/10.1086/517897*.

Eysenbach, Gunther. 2011. "Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact." *J Med Internet Res*.13(4):e123. *https://doi.org/10.2196/jmir.2012*.

Falagas, Matthew, and Vangelis Alexiou. 2008. "The Top-Ten in Journal Impact Factor Manipulation." *Archivum Immunologiae et Therapiae Experimentalis* 56:223–226. *https://doi.org/10.1007/s00005-008-0024-5*.

Franck, Georg. 1999. "Scientific Communication—A Vanity Fair?" *Science* 286:53–55. *https://doi.org/10.1126/science.286.5437.53*.

Goodhart, Charles. 1981. "Problems of Monetary Management: The U.K. Experience." Anthony S. Courakis (ed.), *Inflation, Depression, and Economic Policy in the West.* (Rowman & Littlefield): 111–146. *https://doi.org/10.1007/978-1-349-17295-5_4*.

Gregg, Gordon, Jennifer Lin, Richard Cave, and Ralph Dandrea. 2015. "The Question of Data Integrity in Article-Level Metrics." *PLOS Biology* 13:e1002161. *https://doi.org/10.1371/journal.pbio.1002161*.

Hill, Steven, 2015. Higher Education Funding Council blog. Published July 27, 2015. Accessed June 10, 2017. *http://blog.hefce.ac.uk/2015/07/27/dimensions-of-quality-research/*.

Holmberg, Kim. 2014. "How Concerned Should We Be About Gaming Altmetrics?" *Social Science Research blog*, July 11. Accessed May 30, 2016. *http://www.socialsciencespace.com/2014/07/altmetrics-may-be-able-to-help-in-evaluating-societal-reach-but-research-significance-must-be-peer-reviewed/*.

Holmberg, Kim. 2015. Altmetrics for Information Professionals. *http://www.sciencedirect.com/science/book/9780081002735*.

Lin, Jennifer. 2012. "A Case Study in Anti-Gaming Mechanisms for Altmetrics: PLoS ALMs and DataTrust." *ACM Web Science Conference 2012 Workshop*. Evanston, IL. Accessed May 29, 2016. *http://altmetrics.org/altmetrics12/lin/*.

Lin, Jennifer, Fiona Murphy, Michael Taylor, and Liz Allen. 2017. "Building the Infrastructure to Make Science Metrics More Scientific." *F1000Research* 5:2897. *https://doi.org/10.12688/f1000research.10422.2*. *http://blog.crossref.org/2015/09/det-poised-for-launch.html*.

Lin, Jennifer, and Martin Fenner. 2013. "Altmetrics in Evolution: Defining & Redefining the Ontology of Article-Level Metrics." *Information Standards Quarterly* 25(2):20–26. *https://doi.org//10.3789/isqv25no2.2013.04*.

Liu, Chun Li, Yue Quan Xu, Hui Wu, Si Si Chen, and Ji Jun Guo. 2013. "Correlation and Interaction Visualization of Altmetric Indicators Extracted from Scholarly Social Network Activities: Dimensions and Structure." *Journal of Medical Internet Research* 15(11):e259. *https://doi.org/10.2196/jmir.2707*.

McVeigh, Marie. 2002. "Journal Self-Citation in the Journal Citation Reports." *Thompson Reuters White Papers*. Accessed June 1, 2016. *http://wokinfo.com/essays/journal-self-citation-jcr/*.

Moore, Samuel, Cameron Neylon, Martin Paul Eve, Daniel Paul O'Donnell, and Damian Pattinson. 2017. "'Excellence R Us': University Research and the Fetishisation of Excellence." *Palgrave Communications* 3, Article number: 16105. *https://doi.org/0.1057/palcomms.2016.105*. *http://wokinfo.com/essays/journal-self-citation-jcr/*.

Neylon, Cameron. 2014. "PLOS Submission to HEFCE RFI on Metrics in Research Assessment." *figshare*. *https://doi.org/10.6084/m9.figshare.1089555.v1*. *https://dx.doi.org/10.6084/m9.figshare.1089555.v1*.

NISO. *RP-25–201x-3, Altmetrics Data Quality Code of Conduct—Draft for Public Comment*. Accessed June 19, 2017. *http://www.niso.org/apps/group_public/document.php?document_id=16121&wg_abbrev=altmetrics-quality*.

NISO. 2014. *NISO Altmetrics Initiative Phase 1 White Paper*. Accessed June 19, 2017. *http://www.niso.org/apps/group_public/document.php?document_id=13809&wg_abbrev=altmetrics*.

Opatrny, Tomas. 2008. "Playing the System to Give Low-Impact Journal More Clout." *Nature* 455:7210. *https://doi.org/10.1038/455167b*.

PLOS. 2012. "ALM Workshop 2012 Report." *Figshare*. *https://doi.org/10.6084/m9.figshare.98828.v2*.

PLOS Medicine Editors. 2006. "The Impact Factor Game." *PLOS Med* 3(6):e291. *https://doi.org/10.1371/journal.pmed.0030291*.

Power, Michael, Tobias Scheytt, Kim Soin, and Kerstin Sahlin. 2009. "Reputational Risk as a Logic of Organizing in Late Modernity." *Organization Studies* 30:301–324. *https://doi.org/10.1177/0170840608101482*.

Priem J., P. Groth, and D. Taraborelli. 2012. "The Altmetrics Collection." *PLOS One* 7(11):e48753. *https://doi.org/10.1371/journal.pone.0048753*.

Priem, J., D.Taraborelli, P. Groth, and C. Neylon. 2010. *Altmetrics: A Manifesto*. Accessed May 28, 2016. *http://altmetrics.org/manifesto*.

Priem, Jason, Heather Piwowar, and Brad Hemminger. 2012. "Altmetrics in the Wild: An Exploratory Study of Impact Metrics Based on Social Media." *arXiv*. arXiv:1203.4745.

Shuai, X., A. Pepe, and J. Bollen. 2012. "How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations." *PLOS One* 7(11):e47523. *https://doi.org/10.1371/journal.pone.0047523*.

Strathern, Marilyn. 2000. *Audit Culture*. New York: Routledge.

Thelwall, Mike, Stephanie Haustein, Vincent Larivière, and Cassidy Sugimoto. 2013. "Do Altmetrics Work? Twitter and Ten Other Social Web Services." *PLOS One* 8(5):e64841. *https://doi.org/10.1371/journal.pone.0064841*.

Tuchman, Gaye. 2012. "Commodifying the Academic Self." *Inside Higher Ed*, February 6. Accessed May 30, 2016. *https://www.insidehighered.com/views/2012/02/06/essay-gaming-citation-index-measures*.

van Noorden, Richard. 2013. "Brazilian Citation Scheme Outed." *Nature News* 500:510–511. *https://doi.org/10.1038/500510a*.

Wilhite, A., and E. Fong. 2012. "Coercive Citation in Academic Publishing." *Science* 335(6068):532–543. *https://doi.org/10.1126/science.1212540*.

Wilsdon, James, Liz Allen, Eleonora Belfiore, Philip Campbell, Stephen Curry, Seven Hill, Richard Jones, Roger Kain, Simon Kerridge, Mike Thelwall, Jane

Tinkler, Ian Viney, Paul Wouters, Jude Hill, and Ben Johnson. 2015. *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. https://doi.org/10.13140/RG.2.1.4929.1363*.

Zahedi, Zohreh, Rodrigo Costas, and Paul Wouters. 2014. "How Well Developed Are Altmetrics? A Cross-Disciplinary Analysis of the Presence of 'Alternative Metrics' in Scientific Publications." *Scientometrics* 101:1491–1513. *https://doi.org/10.1007/s11192-014-1264-0*.

17

Why We Could Stop Worrying About Gaming Metrics If We Stopped Using Journal Articles for Publishing Scientific Research

Elizabeth Wager

The problem of "gaming" metrics arises because journal articles are used both to report research and to measure the productivity of researchers. Since publications in peer-reviewed journals are viewed as the "currency" of academia, we have created an incentive for researchers to cheat. This cheating ranges from questionable practices such as "salami-sliced" publications, redundant publications, and gift authorship to serious misconduct such as data fabrication and plagiarism. Pressure to publish in "high-impact" journals also creates incentives for data falsification and misleading reporting, since these highly selective journals seek unusual and interesting findings. (This may be one reason why retraction rates have been shown to correlate with Journal Impact Factors [Fang and Casadevall, 2011].)

One radical solution to the problem of such metrics gaming is to stop publishing scientific research in the form of traditional journal articles and to develop new methods both for disseminating findings and for measuring research productivity. For research that generates numeric or digital data, abandoning journal articles as the medium for disseminating primary results would have several benefits. This chapter uses the reporting of clinical trials to consider and exemplify the problems with traditional journal articles and the benefits of abandoning them in favor of alternative methods of dissemination.

A paradox of medical publication is that, while researchers report being under intense pressure to publish, nevertheless many clinical trials are never published. This causes the medical literature to be seriously skewed. Trials that fail to produce a statistically significant result, or that give results that are disappointing to the sponsor or investigator, are the most likely not to be published. This causes serious publication bias, making the published evidence base unreliable. Estimates of the extent of nonpublication vary, but several studies have suggested that as many as half of all clinical trials are not published (Song et al., 2010; Chan et al., 2014).

Reasons for nonpublication vary but the difficulties of preparing a journal article and then of getting it accepted in a "good" journal undoubtedly contribute (Smyth et al., 2011). Although there are now several journals that explicitly welcome reports of small studies, replication experiments, and nonsignificant findings, academics either continue to believe that such work is unlikely to be accepted, or fear that publishing it may harm their reputation or career. Pressure from sponsors to suppress inconvenient findings and lack of incentives to publish all trials are probably also factors. Current systems of measuring academic productivity focus almost exclusively on the publication of journal articles, often with an emphasis on high-impact journals. Researchers may therefore be incentivized to abandon disappointing avenues of research, leaving them unpublished, to focus on new work, which they hope will be more successful. Competition between researchers may also create disincentives for reporting failed methods.

Another problem with traditional publication models is that many trials are only partially published. While a three-thousand-word article may be convenient for many readers, the space constraints imposed mean that many details of clinical trials are not included in journal articles. Deficiencies in describing both the methods and the results have been clearly documented and are worryingly prevalent (Hoffmann et al., 2014). Missing details about the methods prevent findings from being replicated, trial quality from being properly assessed, and new techniques from being implemented. Partial reporting of results may contribute to publication bias since statistically significant or "positive" findings are more likely to be included than inconclusive or inconvenient ones (Chan et al., 2004). Incomplete reporting (whether deliberate or unintentional) can rarely be detected from journal articles alone and cannot be spotted during peer review unless reviewers have access to, and are prepared to carefully check, the protocol.

The lack of linkage between study protocols, underlying data, and research reports also creates the possibility for other problems. Peer review of a journal article is often done without reference to the study protocol and data analysis plan, which describe the original design for the research and are therefore essential for interpreting the findings. Comparisons of protocols and journal articles have shown that a worryingly high proportion of published articles (around sixty percent according to some studies) do not report the primary outcome specified in the original study design, or switch primary and secondary outcomes (Chan et al., 2004). This undermines the reliability of the analysis and may reduce the statistical power of the analysis. Linking documents such as protocols, investigator brochures, patient information leaflets, and the various presentations of findings, including clinical trial reports prepared for regulators, conference presentations, results postings on trial registers, and descriptive articles, would not only be more efficient, but also produce a more reliable account of the research (Glasziou et al., 2014). Such linkage is technically possible and may be partially automated, but could be achieved more effectively if supported by funders and regulators (Goldacre and Gray, 2016).

Another weakness of traditional journal articles, and the way in which they are peer reviewed, is that this system provides little or no safeguards against incorrect statistical analysis. Such problems have been well documented in published articles even in the top-ranking journals and may occur deliberately, to emphasize or mask effects, or inadvertently through researcher ignorance (Altman, 2002). Inappropriate handling of missing data or outliers can also affect outcomes. Unless reviewers have access to the raw data, and sufficient statistical expertise, this is usually not detected by peer review. The traditional journal article, which shows only the analyzed aggregate data, rarely reveals such problems.

While inappropriate statistical methods appear to be relatively common, deliberate research fraud is probably much rarer, but not so rare that it can safely be ignored. However, sophisticated data falsification or fabrication is very hard to detect from traditional publications (Carlisle, 2012). Reviewers and readers see only the final results of analyses and have no access to the underlying data. Falsification of digital images may be detected if journals have access to original images and screen them, but this is time consuming (Linkert, 2010). Therefore, image manipulation is often detected only after publication by readers. Several cases of fraudulent image manipulation leading to retractions have been identified in this way (e.g., via alerts on PubPeer or directly to the journal).

One explanation given for the longevity of the journal article, which has changed little in 350 years in terms of length or format (Wager, 2006), is that readers like them. However, research articles are read for a variety of purposes, and one size does not necessarily suit all users (Altman, 2015). Many readers prefer a short summary, and journals have responded by including abstracts or even shortening the format (e.g., the *BMJ's* "Pico" format [Jain, 2014]). However, other readers may seek more detail in the methods, so that they can replicate the findings or use the technique.

Articles are also used to create systematic reviews, which may combine findings using the statistical techniques of meta-analysis. This is often hampered by deficiencies in journal articles, so those compiling the review are forced to contact the authors to seek further information, which is not always forthcoming (Wager, 2006). Also, despite the familiarity with the journal article format, and even after technical editing, journal articles generally remain hard to read (Wager and Middleton, 2002).

Despite the predominance of the journal article, research is usually reported in several formats, such as a conference abstract, journal article, and press release and, increasingly, on trial registers. Other reports may also be prepared but shared only with regulators. Relevant information about the research may also be contained in the trial protocol and other documents such as the investigator brochure and participant information (Chalmers and Altman, 1999). Not only is the production of these multiple formats inefficient, but because versions are not linked, they may be inconsistent (Francis et al., 2013; Glasziou et al., 2014). In such cases, it may be unclear which is the correct version. Discrepancies may be due to simple errors or more complex factors such as different methods of data handling and analysis required by regulatory authorities and journal peer reviewers. The reporting of adverse events is especially prone to this (e.g., regulatory reports may include all adverse events while journal articles may report only those that were considered likely to have been caused by the treatment; classification of adverse events by severity may also vary) (Hughes et al., 2014). Once again, alternative publication models, especially those allowing linkage to the underlying data, might reduce these problems.

Conclusions

Familiarity with the conventional format of journal articles may lull users into a false sense of assurance. Yet, despite its familiarity, this format has many deficiencies as mentioned above and noted by several previous commentators (Smith, 1992; Chalmers and Altman, 1999; Smith and Roberts, 2006; Altman, 2014; Tracz, 2015). Furthermore, because journal articles are the "currency" of academia, and highly profitable for publishers, there is built-in resistance to change. The barriers to radically rethinking the way in which we disseminate research results are more social and cultural than technological.

The ideal system for reporting research would link the underlying data, appropriately labeled, with the full methods in the protocol and the entry on the trial register (Chalmers and Altman, 1999; Glasziou et al., 2014). Text elements should probably be highly structured and machine readable (Altman, 2015). This would ensure completeness and consistency. The use of structured fields and online reports would mean that many checking functions could be automated (Wager, 2006). Structuring and linkage of information should reduce the burden of producing different formats and increase consistency. The resulting "report" might appear dull (at least to human readers), but that seems a small price to pay for accuracy. Incentives should be shifted away from publishing articles in journals toward systems that ensure that all research is publicly posted (for example, by funders withholding the final part of grants until this has been done). This radically different method of reporting research would also bring opportunities for new quality control mechanisms that would, most likely, replace traditional prepublication peer review, which has also been shown to have serious shortcomings and is not well suited to reviewing datasets (Wager and Jefferson, 2001; Tracz, 2015).

If scholarly journals were no longer the medium for publishing primary research findings, journal articles would cease to be the "currency" by which research output was measured. This would give an opportunity to develop better metrics. There are already proposals of new systems to give credit for data "authorship" (Bierer et al., 2017) and best practice for publishing raw clinical trial data (Hrynaszkiewicz and Altman, 2009). While structured reports and raw data might be highly efficient methods for disseminating research findings, there would still be a role for commentary, interpretation, and synthesis, all of which could be provided by scholarly journals.

Viewed in this way, the deficiencies of the current system for publishing research seem enormous, and the potential benefits from radical change seem obvious. Nevertheless, change has been technically possible and clearly proposed for almost twenty years. At the moment, it remains unclear whether dissatisfaction with journal articles might kill the current system of metrics, or whether concerns about gaming and unfair metrics could kill the journal article.

References

Altman, Douglas G. 2002. "Poor-Quality Medical Research. What Can Journals Do?" *JAMA* 287:2765–2767.

Altman, Douglas G. 2015. "Making Research Articles Fit for Purpose: Structured Reporting of Key Methods and Findings." *Trials* 16:53.

Bierer, Barbara E., Mercè Crosas, and Heather H. Pierce. 2017. "Data Authorship as an Incentive to Data Sharing." *New England Journal of Medicine* 376:1684–1687.

Carlisle, John. 2012. “The Analysis of 168 Randomised Controlled Trials to Test Data Integrity.” *Anaesthesia* 67:521–537.

Chalmers Ian, and Douglas G. Altman. 1999. “How Can Medical Journals Help Prevent Poor Medical Research? Some Opportunities Presented by Electronic Publishing.” *Lancet* 353:490–493.

Chan, An-Wen, Asbjørn Hróbjartsson, Mette T. Haahr, Peter C. Gøtzsche, and Douglas G. Altman. 2004. “Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles.” *JAMA* 291:2457–2465.

Chan, An-Wen, Fujian Song, Andrew Vickers, Tom Jefferson, Kay Dickersin, Peter C. Gøtzsche, Harlan M. Krumholz, Davina Ghersi, and H. Bart van der Worp. 2014. “Increasing Value and Reducing Waste: Addressing Inaccessible Research.” *The Lancet* 383:257–266.

Fang, Ferric C., and Arturo Casadevall. 2011. “Retracted Science and the Retraction Index.” *Infection and Immunity* 79:3855–3859.

Francis, Darrel P., Michael Mielewczik, David Zargaran, and Graham D. Cole. 2013. “Autologous Bone Marrow-Derived Stem Cell Therapy in Heart Disease: Discrepancies and Contradictions.” *International Journal of Cardiology* 168:3381–3403.

Goldacre, Ben, and Jonathan Gray. “OpenTrials: Towards a Collaborative Open Database of All Available Information on All Clinical Trials.” *Trials* 17:164.

Glasziou, Paul, Douglas G. Altman, Patrick Bossuyt, Isabelle Boutron, Mike Clarke, Steven Julious, Susan Michie, David Moher, and Elizabeth Wager. 2014. “Reducing Waste from Incomplete or Unusable Reports of Biomedical Research.” *The Lancet* 383:267–276.

Hoffmann, Tammy C., Chrissy Erueti, and Paul P. Glasziou. 2013. “Poor Description of Non-Pharmacological Interventions: Analysis of Consecutive Sample of Randomised Trials.” *BMJ* 347:f3755.

Hrynaszkiewicz, Iain, and Douglas G. Altman. 2009. “Towards Agreement on Best Practice for Publishing Raw Clinical Trial Data.” *Trials* 10:17.

Hughes, Shannon, David Cohen, and Rachel Jaggi. 2014. “Differences in Reporting Serious Adverse Events in Industry Sponsored Clinical Trial Registries and Journal Articles on Antidepressant and Antipsychotic Drugs: A Cross-Sectional Study.” *BMJ Open* 4:e005535.

Jain, Anita. 2014. “Evolving Trends in Research Communication in a Digital Age: Examples from the BMJ.” *Keio Journal of Medicine* 63:67–68.

Linkert, M., C. T. Rueden, C. Allan, J. M. Burel, W. Moore, A. Patterson, B. Loranger, J. Moore, C. Neves, D. Macdonald, A. Tarkowska, C. Sticco, E. Hill, M. Rossner, K. W. Eliceiri, and J. R. Swedlow. 2010. “Metadata Matters: Access to Image Data in the Real World.” *Journal of Cell Biology* 189:777–782.

Smith, Richard. 1992. “The End of Scientific Journals?” *BMJ* 28:304.

Smith, Richard, and Ian Roberts. 2006. “Patient Safety Requires a New Way to Publish Clinical Trials.” *PLoS Clinical Trials* 2006;1:e6.

Smyth, R. M. D., J. J. Kirkham, A. Jacoby, D. G. Altman, C. Gamble, and P. R. Williamson. 2011. "Frequency and Reasons for Outcome Reporting Bias in Clinical Trials: Interviews with Trialists." *BMJ* 342:c7153.

Song, Fujian, S. Parekh, L. Hooper, Y. K. Loke, J. Ryder, A. J. Sutton, C. Hing, C. S. Kwok, C. Pang, and I. Harvey. 2010. "Dissemination and Publication of Research Findings: An Updated Review of Related Biases." *Health Technology Assessment* 2010; Vol. 14: No. 8. http://www.journalslibrary.nihr.ac.uk/__data/assets/pdf_file/0005/64751/FullReport-hta14080.pdf. Accessed February 29, 2016.

Tracz, Vitek. 2015. "The Five Deadly Sins of Science Publishing." *F1000 Research* 4:112.

Wager, Elizabeth. 2006. "Publishing Clinical Trial Results: The Future Beckons." *PLoS Clinical Trials* 1:e31.

Wager, Elizabeth, and Philippa Middleton. 2002. "Effects of Technical Editing in Biomedical Journals: A Systematic Review." *JAMA* 287:2821–2824.

Wager, Elizabeth, and Tom Jefferson. 2001. "Shortcomings of Peer Review in Biomedical Journals." *Learned Publishing* 14:257–263.

IV

Mimicry for Parody or Profit

Mimicry for Parody or Profit looks at fakeries (and critiques thereof) that are rooted in "brand appropriation," parody, and hoax. For instance, fake universities that sell degrees without any attempt at educating (not even online) tend to assume Ivy League–sounding names. What we see are attempts to mimic a "brand" (of a university or a journal), down to the look of their websites, rather than just copying or pirating a specific product (like an article). (See also the mimicking of journals in Ike Antkare's chapter above). It is only fitting that some critiques of these spammish mimicries are humorous mimicries themselves, like submitting computer-generated articles that mimic the academic genre, creating fake scientists with funny names to author fake articles to make fun of the arbitrariness of citation evidence, or joining spammish editorial boards under comic aliases to expose them.

Drawing from his research on the cultural history of spam, Finn Brunton asks why some new journals that mimic well-established journals are called "spam journals," while also analyzing how the "spammishness" of so-called spam journals is different from classic spam. Looking at both the beneficiaries and the victims of spam journals, Brunton suggests that—as with other forms of spam—the answer is more complex than it seems. Are the junior professors, the overworked adjuncts, or the scholars from resource-poor universities who publish in spam journals victims or happy customers? Similarly, are the established publishers from the global North to be taken at face value when they accuse these "spam" journals of discrediting the whole system of scholarly publishing (and their "good" journals in it)? Or are they instead benefiting from the existence of these "spam" venues that allow them to construe their journals as "good" by simply contrasting them with the so-called spam ones? The answer is largely in the eye of the beholder, whether one falls (or pretends to fall) for the mimicry, or rides with the mimicry to create the effect of an original.

Marie-Andrée Jacob too dwells on the constitutive tension between the original and the counterfeit: “What is most crucial to recognize, here, is the inevitability rather than exceptionality of the eruption of counterfeit scientific journals. It is unhelpful to see the ‘make-believe’ as anomalous.” Unlike Brunton, however, she does not pursue the *cui bono?* question as a window on how the relation between the original and the copy can be construed as either one of opposition or identity. To her, the original and the counterfeit are always already mutually constitutive and they can be either collapsed or teased apart only through much labor and dexterity. Even a publication misconduct watchdog like COPE has to watch out for the appropriation of its trademark.

Alessandro Delfanti shows how mimicry in academic publishing does not need to be driven by a desire for profit, or by an attempt to emulate, but may be adopted as a critical, even subversive, gesture. His example concerns viXra.org, a new fully open science preprint server whose name is the mirror image of arXiv.org, the almost legendary preprint server housed at Cornell. (The layout of the viXra.org site is also virtually identical, minus the color palette and the Cornell University logo, to the arXiv.org site). But unlike the newly developed bioaRxiv.org server aimed at providing the life sciences with a service comparable to what physicists and mathematicians had in arXiv.org—a partial “mimicry” that bioRxiv meant as gesture of appreciation of its older “ancestor”—viXra.org wants to “shadow” arXiv.org to make a pointed critical statement: “The visual design of viXra.org (but not its content) is a parody of arXiv.org to highlight Cornell University’s unacceptable censorship policy. ViXra is also an experiment to see what kind of scientific work is being excluded by the arXiv.”[1] Obviously, viXra does not wish to mislead authors into uploading papers on their site believing that it is arXiv’s, but rather to make a statement that viXra is the “good” server by virtue of being the reverse mirror image of arXiv.org, which they criticize for having introduced vetting practices that exclude amateur scientists and other authors who do not happen to have a “proper” institutional affiliation. Mimicry, in this case, expresses a distinctly ad hominem criticism, a parody so specific to its target that ends up assuming its same (albeit reverse) look. (One could say that viXra participates in the carnivalesque discourse of inversion in the sense that it presents itself, literally, as arXiv upside-down.)

Alexandra Lippman explores the carnivalesque in the ways in which the watchdogs and pranksters mock spammish journals. Seeking a laugh, these pranksters create academic paper generators, humorous blogs, hoaxes, and pseudonymous personae to reveal the lack—or poor quality—of review on the part of what they see as predatory journals and conferences. This, however, is a rather unusual kind of carnivalesque that, unlike the traditional form studied by Mikhail Bakhtin, does not target authority but rather pokes fun at the fraudsters.

Notes

1. http://vixra.org/why.

18

Making People and Influencing Friends: Citation Networks and the Appearance of Significance

Finn Brunton

Like many academics, I became interested in predatory and "spam" journals and new forms of academic misconduct when I started receiving the invitations to submit an article, any article—fast turnaround! low prices!—or present at a conference, review for a journal, or even join their board. My initial amusement at being asked to submit something about immunology or linguistics (a testament to the interdisciplinary character of modern university life, perhaps) was in retrospect a stage of professionalization through which every academic with an email address now passes: being a potential mark for a family of new, related scams. I had more than an accidental interest in this, however. I had written a book about the history of spam, and the question of what made these journals *spam* journals was something I wanted to investigate (Brunton, 2013). They are both like and unlike other forms of spam—as Mario Biagioli puts it in the introduction to this volume and Alexandra Lippman explains in her chapter, they are neither spam nor not spam, but *spammish*, and part of larger systems of imitation, collusion, and gaming metrics. I will discuss both parallels and divergences between spam itself and academic platforms dubbed with that title, and present you with two closely related lessons from spam for studying academic metrics and their abuse. The first has to do with how we study the secondary markets around misconduct and the second with understanding the larger networks that benefit from the work of platforms for scholarly misbehavior. Together, I hope these will enrich our picture of what academia is now in danger of becoming, through how it can be *gamed*.

First, though, why is it that these journals are called "spam journals"? There's an obvious answer to this question, and a deeper answer with more to tell us. The obvious answer is straightforward, but still worth dwelling on for clarity: these are journals that most of us encounter through the delivery channel of unrequested and indiscriminate email. Indeed, these journals replicate many classic spam strategies in the content of their messages and their web presence—working off the same well-established playbook used by stock touts and boiler-room investment operations, deadstock salesmen and off-brand pharmacists, pornographers and confidence men.

Their messages feature the same array of rhetorical appeals and come-ons, though geared toward an academic audience: lower fees, faster publishing cycles, sympathetic reviewers, more favorable impact metrics. (Given what is about to be discussed, the reader should add quotes liberally: favorable "impact metrics," sympathetic "reviewers." The only solidly realistic thing is the fees.) Like phishing emails—which try to trick recipients into mistaking an illegitimate message for one from Facebook or their bank by aping the style, design, and markup—many of these messages have the boxy, early-2000s design particular to more reputable journals, sometimes in more or less direct reference. (Of course, we also see cases of *actual* academic journal phishing, with scammers replicating the title, text, and sometimes look and feel of a reputable journal's site so unsuspecting scholars will submit papers, with fees, for review.[1]) They use US and European post office boxes as mail drops, to produce an appearance of legitimacy, much as the 1990s spammers would have business addresses in the signature of their messages, to give the impression of an actual office somewhere, an institutional relationship, the possibility of redress—although with the distinction that the spam journal drops often actually *exist,* to redirect mailed checks.

The same strategy applies to the use of evasive and misleading contact information and lists of personnel, and links to legitimate venues as an indirect assertion of bona fides—as advance-fee fraud spammers would include links to reliable news organizations and trustworthy banks to imply their trustworthiness in turn. Burkhard Morgenstern's chapter in this book shows the recruitment process for getting real, legitimate names on illegitimate organizations at its most egregious; some groups avoid the difficulty in gathering reputable editors by featuring an editorial board of wholly fictional people, like the *East European Scientific Journal* (Beall, 2016). Spammers, whether selling weight-loss pills and cadging credit card numbers or requesting scholarly papers in marine biology, always need to minimize the effort involved in producing content like email text or a site's landing page—the price of a business built on casting the widest possible net—and therefore reuse existing content to rapidly generate new venues for potential customers. The *American Journal of Pure and Applied*

Mathematics is an "international journal of high quality" with an astonishing range of interests (from statistics and probability to number theory, wavelets, and Banach spaces), which it shares, word for word, typos included, with the *Global Journal of Pure and Applied Mathematics*.[2]

All of these are classic spam strategies, and immediately explain how attempts to get junior scholars to submit papers to generic-sounding journals can be understood with the same word used for emails promising to restore virility. However, there is a deeper connection between these two things—between bottom-feeding ad scams and the *Global Journal of Pure and Applied Mathematics*. This is the second, subtler answer: both have to do with producing the appearance of salience.

Following close on the heels of the development of the first search engines—and in some cases driving their innovation—was the spammer's project of generating this appearance. A spammer wants to get a link or some content in front of potential victims, through whatever platform the victim employs. Therefore, the spammer needs to convince the evaluative metrics that are judging the importance of the information. The content cannot actually stand up to scrutiny on its own—and certainly not human scrutiny, no more than the gibberish papers, screeds, and obvious pseudoscience published by spam journals could hold up for an actual scholarly audience. To appear to satisfy whatever criterion is being analyzed—relevance or utility or popularity—the spammer needs to generate populations of "users" whose linking and citation behavior looks, from the right distance and by the right metric, like a popular endorsement.

This is actually a problem that originates with scholarly and legal citation networks (the two groups of people most obsessed with citation are academics and professional spammers) because those networks were the inspiration for Google's ranking algorithm, which determines the salience of a web page for a given query. "The PageRank Citation Ranking: Bringing Order to the Web," the paper that outlined the system at the core of Google's process, is about adapting the citation-analysis model from one platform to the other, though "there are a number of significant differences between web pages and academic publications" (Page et al., 1999). Given our subject in this book, the most important difference they observe offers us a painful example of historical irony: "Unlike academic papers which are scrupulously reviewed, web pages proliferate free of quality control or publishing costs…[H]uge numbers of pages can be created easily, artificially inflating citation counts." We know that this paper is important, because "The PageRank Citation Ranking" is itself one of the most cited papers in computer science; a similar process can help us determine that one site is more significant than another (combined with many other subassemblies of algorithms and analytic tools).

The PageRank system, developed in part to counteract the first generations of lexical spam built to fool keyword-based search engines, set off a still-ongoing arms race. Automating the academic process—treating links as cites for determining the really important papers, as it were—worked well, but new platforms enabled the rapid production of venues (web pages, wikis, blogs) that could produce huge volumes of citations: linkfarms of hosted blogs, spam entries added to hacked wikis, automated comments, and so on, up to contemporary platforms like Twitter and "likefarming" on Facebook. Of course, the preventative strategies have likewise evolved, taking approaches like analyzing the social network graph as a whole, rather than the content of individual posts, to spot bad actors. The telling characteristic of early linkfarms and likefarms was that they were densely interlinked but lonely: that is, they would all link to each other but nobody outside their group would ever link to them.[3] That made them relatively easy to identify and eliminate. Spammer strategies evolved in response: spam Twitter accounts have a mix of fake and real followers, for instance, many of them in deliberately small numbers. The methods have changed, but the goal remains the same—making people, and influencing "friends," in the eyes of the machines.

That last detail is a crucial one, and brings us back to predatory journals and other new forms of academic misconduct. Spam Twitter accounts, blogs, and web pages are seldom meant to be seen by human beings; their speciousness is obvious to even the casual reader. They exist to influence metrics—follower counts, likes and retweets, views of videos, listens of songs, bumping up search engine rankings—with influence on people as a secondary or tertiary effect. The use of "private blog networks" (PBNs) in the search engine spamming community, networks of huge numbers of automatically generated blogs that exist to link to other sites rather than be encountered by humans, is the equivalent of a dodgy academic paper deliberately published in a spam journal.[4] The paper was never meant to be *read;* it exists to appear as a line in a CV, one among many, to act as a token in an assessment process.

Here, we can see the outline of the first lesson from spam for studying these new forms of misconduct: to pay attention to the *secondary markets*. When we think of spam we tend to think of the specific examples we happen to encounter—an email, a Twitter account, a blog stitched together from hundreds of cut-up public domain sources—but those are only the outlying, human-facing parts of the business, whereas whole secondary economies and marketplaces exist to build the tools the spammers need, marketplaces that make deeper trends visible. To get rich in a gold rush, you don't go prospecting: you sell shovels, assays, work pants. Spammers rely on many kinds of back-end infrastructure, on payment systems and accomplice banks—and, like academics, on metrics companies and impact measuring tools. Quick! Is this promise from a spam journal company, or a spammy search engine manipulation company: "Citation Flow is based on stronger, iterative mathematical logical than the old metric of ACRank."[5] (Odd sentence structure in original.) What about this one? "The Impact Factor is calculated by several scientific methods including citation anlaysis [*sic*]. No Evaluation Processing Fees."[6] As Michael Power argues elsewhere in this book, the business of analytics and metrics starts to drive the product to be measured, rather than the other way around: the outcome becomes the target. Spammers and academics alike will look at what has impact—what can deliver good metrics—and tailor their spam campaigns and search engine optimization, or research projects, accordingly.

The proliferation of predatory journals, like the proliferation of spam email, sites, and accounts, includes the development of secondary markets and facilitation tools: new kinds of misleading or fraudulent metrics and document and object identifiers—the heralds of an entire parallel scholarly apparatus, a crooked ancillary economy.[7] Keeping a close eye on those will reveal, as it has with spam, deeper trends in this domain beyond any individual misconduct. It is on that point that the second lesson from spam rests.

In the same way that we tend to think of spam as "this email"—and of spam publishing as "this fraudulent journal"—when there is much more to see in the secondary layers of infrastructure and facilitation, we tend to think of spam in terms of straightforward and singular victims. Of course there are people victimized by phishing messages, identity theft, and bogus products, but aside from those cases, the question of victims becomes more complex. Take spam Twitter accounts, which are simultaneously a misuse of Twitter but also somewhat to Twitter's benefit. A spambot population artificially inflates their user numbers and metrics of activity and gives their users, both by accident and by design, similarly inflated follower counts. People who thought that they were wildly popular on some legitimate grounds, or wanted to be seen so, are suddenly revealed in their networked insignificance during a legitimate cleanup effort, often to their outrage.[8] In these moments of exposure—when both deliberately and accidentally pumped-up follower counts are exposed, we can see that social networks have undergone a process akin to the shift of scientific papers from units of knowledge to "accounting units" described by Yves Gingras in this book. What began—at least notionally—as relationships between people has become an accounting unit for measuring significance through retweets, responses, likes, and "followers," in ways that benefit, as well as harm, the platforms and the victims alike. This is only one of the confusing *larger networks of benefit at play* in spam. This is not to say that spam is actually a good thing, but simply that we do not necessarily have an accurate picture of who benefits and who the victims are, which can explain some puzzles about how it works.

Spam journals and "predatory" publishing systems have more complex answers to the *cui bono?* question than we may at first assume. Who is preyed on by the predatory journal? Sometimes a junior academic who doesn't realize that they're being taken advantage of. But every scholar wants another line on their CV, and so does their department and their administration—which perhaps gets rewarded by institutional assessment tools—and even their country, seeking to reward "performance" without having a nuanced picture of what precisely performance is, for any given discipline, and how it should be measured. Spam journals make it possible for chronically overworked adjunct faculty to keep up a brisk publishing pace, for people without significant academic resources at their home university to rack up impressive records of "international" activity. For, let us say, the equivalent of $90 US? A fair price for services rendered, perhaps: no one has to read the resulting paper, no library needs to subscribe, and impact is automatically generated by a more-or-less imaginary system.

Additionally, as spam helped to put "legitimate" advertising into a better light on the old, resolutely noncommercial internet (we may be sending you ads, but at least we're not like the diet pills, porn, and malware crew), spam journals indirectly valorize the seriousness and status of mostly developed-world, Global North journals—the kinds of venues in which *respectable* scholars publish. De Rijcke and Stöckelová put this best in their chapter: "predatory publishing and its concomitant practices are not outside of the research system but emerge at the heart of them and are embedded within them," reinforcing the distance between the "'international' West or North on the one hand, and a 'parochial' East or South on the other." The top-tier journals and the major universities already constitute a kind of *de facto* citation cartel, colluding in a shared economy of cultural, social, and scientific capital with their status further reinforced by all these incompetent, ersatz rip-offs: the genuine article, accept no substitutes! (Indeed, Marie-Andrée Jacob argues in the following chapter that counterfeits and originals are inextricably linked: “counterfeiting solidifies the ‘template’ of elite science and keeps it intact.”)

This is not an exercise in devil’s advocacy, to defend these pseudoscience-publishing, author-exploiting, corrupt, open-access-trashing journals, but a chance to think about who gets something from their existence, and what contributes, directly and indirectly, to the environment in which they thrive. Of what larger networks of value are they a part? What are the secondary markets and ancillary products that profit from them? As spam explains some otherwise enigmatic developments in the history of the internet, spam journals give us a bleak, oblique portrait of how academia is being measured and evaluated now—and what it is in danger of becoming.

Notes

1. Jeffrey Beall maintained an extensive list of various predatory, problematic, and suspect scholarly publishing projects. He took the site down in early 2017; snapshots of the site are on the Internet Archive. I draw a number of my examples of spam journals from his research and will cite him with Internet Archive links. For instance, a list of these journal phishing projects—which Beall called “hijacked journals”—can be found at https://web.archive.org/web/20160310203111/https://scholarlyoa.com/other-pages/hijacked-journals/. For a specific example, see the Mexican life sciences journal *Ludus Vitalis* (http://www.centrolombardo.edu.mx/ludus-vitalis/) and the phishing site for “Ludus Vitalis” (http://ludusvitalis.org.mx).

2. The identical text can be seen at http://www.academicresearchjournals.com/journal-detail.php?journals_id=37 and http://www.ripublication.com/gjpam.htm. Whether there is a relationship between “Academic Research Journals (India)” and “Research India Publications,” their respective publishers, is left as an exercise for the reader.

3. There has been extensive research on this phenomenon; two exemplary papers are Dennis Fetterly, Mark Manasse, and Marc Najork, “Spam, Damn Spam, and Statistics,” *Proceedings of the 7th WebDB Workshop,* 2004, and Zoltan Gyöngyi and Hector Garcia-Molina, “Link Spam Alliances,” *Proceedings of the 31st VLDB Conference,* 2005.

4. For the interested reader, there are many thorough guides to the particulars of how PBNs are built and operated in the form of various get-started manuals, generally as invitations to pay for services. See, for instance, http://authoritywebsiteincome.com/build-private-blog-network/.

5. “Citation Flow,” Majestic Marketing Search Engine Glossary, https://majestic.com/support/glossary#CitationFlow.

6. “Impact Factor,” CiteFactor Academic Scientific Journals, http://www.citefactor.org/page/impact-factor.

7. These are likewise drawn from Beall's work. See "Misleading Metrics," Scholarly Open Access, https://web.archive.org/web/20160303171335/https://scholarlyoa.com/other-pages/misleading-metrics/.

8. For some broader analysis of the scale of the problem, see Networked Insights, "How Dirty Is Big Data?," March 2015, http://info.networkedinsights.com/Dirty-Data-LP.html. For a very entertaining look at the peculiar structure of Twitter spambot populations (including the data itself), see Terence Eden, "This Is What a Graph of 8,000 Fake Twitter Accounts Looks Like," March 9, 2015, https://shkspr.mobi/blog/2015/03/this-is-what-a-graph-of-8000-fake-twitter-accounts-looks-like/. For a journalistic summary of experiments in buying and detecting spam followers on Twitter, see Daniela Hernandez, "Why Can't Twitter Kill Its Bots?," Fusion, September 21, 2015, http://fusion.net/story/195901/twitter-bots-spam-detection/.

References

Beall, Jeffrey. 2016. "Bogus Polish Journal Has Completely Fake Editorial Board." Scholarly Open Access (blog), February 2, 2016. https://web.archive.org/web/20160213180338/https://scholarlyoa.com/2016/02/02/bogus-polish-journal-has-completely-fake-editorial-board/).

Brunton, Finn. 2013. *Spam: A Shadow History of the Internet.* Cambridge, MA: MIT Press.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. "The PageRank Citation Ranking: Bringing Order to the Web." Stanford InfoLab Technical Report, 1999.

19

Crack Open the Make Believe: Counterfeit, Publication Ethics, and the Global South

Marie-Andrée Jacob

My chapter joins the conversation on metrics and misconduct via the concept of "counterfeit." The research context of my short intervention draws on ethnographic and archival work, engaging the question of how people experience but also imagine legality/illegality. Since 2010, as part of my interest in the category of "publication ethics," I have been conducting ethnographic observations of the quarterly forum of the global charity Committee on Publication Ethics (COPE). My research also looks at how the category of "research misconduct" has taken form in the context of disciplinary adjudication by regulators (Jacob, 2014, 2016a) and of modern patterns of documentation more generally (Jacob, 2017). In brief, I am as much interested if not more in institutional watchdogs of academic misconduct than I am in alleged perpetrators of academic misconduct. Pausing over the mutually exclusive dichotomy of real versus counterfeit journals, my short intervention approaches the idea of counterfeit by way of making three points in relation to public harm and denunciation, the idea of the authentic, and watchdogs. Through these anchor points, I hope we can better see the eruption of counterfeit scientific journals as more inexorable than strange or outrageous. The idea here is not to justify the counterfeit of academic journals by claiming that counterfeit exists elsewhere; it is also not to exoticize or, worse, romanticize counterfeiters. Rather it is to examine it on its own terms, from the point of view of its craft, and to highlight dexterity as one of its most underexplored aspects. As James Siegel has beautifully shown in his ethnography of counterfeiters in contemporary Indonesia (Siegel, 1998), there exists a certain power in making fake university certificates, or fake divorce certificates, and so on. Aside from being about the financial profit it brings, it is a power for crafting "a sort of authority for one's self" or "one's own rubber stamp" and for attesting to one's creative abilities. Given the transformations of scientific research and publishing over the last thirty years, described extensively in the editors' introduction and other chapters to this volume (see Wouters, chapter 4, Gingras, chapter 2, and de Rijcke and Stöckelová, chapter 7, in particular), counterfeit might not be as perplexing as some would like to believe. As I hope to show, it is a rather predicable response, as it is a power that does "make do" and does make things move on for one's self or for others (see Craciun, 2012).

Public Harm and Denunciation

Breaches of research integrity are conceived to have wide-ranging negative consequences for the trustworthiness of science and the health of the public. I do not wish to reiterate or question this view here. By large, my current ethnographic fieldwork on watchdogs of scientific misconduct suggests that this threat of public harm is a key argument that inspires much of the professionalized labor deployed against conduct that gets perceived as incompetent or fraudulent. This threat is assumed rather than demonstrated, but this does not mean that the watchdogs' claims are simple. They are rather complex and sophisticated in their forms, using various registers such advocacy and lobbying, expert discourse, the use of "technologies of integrity verification," vigilantism, and uncovering or hoax to convey their message (Jacob, 2015).

We can unpick the claims of public harm and how these are being deployed in academic misconduct debates through examining the "uncovering" work of watchdogs and journalists targeting so-called "predatory journals." This work exposes the problem of predatory journals in a "public service" style, using a revelatory and denunciatory tone on the basis that if predatory journals are unmasked, they will be less of a threat to science and the public good. One cannot help noticing how this uncovering work is also often performed with humor, and elicits mocking laughter on the part of its audience. The work is meant to ridicule counterfeiters; to inform, but also to make us laugh. As I will explain below, this mocking mode is not innocent, as it automatically grants moral and intellectual superiority to the author of the revelation. Some critiques of this uncovering work see it as a frontal attack on open access, but it also more broadly bashes a scientific subpolity, a subaltern ecosystem within the Global South whose actors attempt to play the metrics game too and do so by mimicking the successful model brand of science.

Take for example the piece of investigative journalism "Who's Afraid of Peer Review" (Bohannon, 2013). The punchy Bohannon article is based on an elaborate hoax concocted by the author, in which a spoof article was submitted and accepted to dozens of open-access journals, thus exposing deficient peer-review practices. The piece bashes the Global South in its explanation of the very making and preparatory work of the sting itself—for instance, an African-sounding name was deliberately used as pseudonym to add credibility to the fake paper.[1] Bohannon's piece also mocks well-established Western scientists from elite institutions who attempt to double-dip, that is, to gain benefits—credit, credentials, lines in the CV—from both the model science and the subpolity of counterfeit journals.

Stings like this expose an alternative ecosystem that has understood very well that one of the most valuable currencies here is precisely what is copiable, what can be slotted in and read into a CV (to be noticed but not necessarily read), and what makes one "make do": the names and the brands of science. Recent research has demonstrated what sorts of currency fake journals produce (Xia et al., 2015). More so, it has debunked assumptions about which "public" or audience is addressed by the fake journals sounding like real journals, and to what extent they harm this public: Jingfeng Xia and his colleagues show that the target audience of these journals is not mainly comprised of readers, users, and stakeholders in the ordinary sense of the words—but rather decision makers within Global South institutions where individual counterfeiters live and hope to "make do," that is, to make a living by keeping their job.

We can illustrate the point further by looking at an analogy from the context of state making. In her study of the make-believe state, Yael Navarro-Yashin (2012) notes that a "wannabe" state has to produce documents to look and act like a state, in other words to perform the state. The entity Navarro-Yashin refers to is not recognized as such under international law: the Turkish Republic of Northern Cyprus (TRNC). Yet tax office, electricity unit, and immigration office documents all carry its logo. These printed logos "do not only represent specific identities and transactions, but also declare legitimacy of the TRNC. They work within the make-believe state, but are not considered legal (and therefore 'real') outside of this self-declared polity" (Navarro-Yashin, 2012). This last sentence points to the currency and leverage of "wannabe" documentation. The TRNC logos echo the point about the effects of counterfeit journals beyond their own polity: counterfeit journals, like the "wannabe" documentation illustrated by Navarro-Yashin, have a more local and affective than large-scale impactful existence. Yet Western Euro-American fears about these journals remain palpable, and the uncovering work helps satisfy the appetite for denunciation, using a form that is easily identifiable to scientists such as the hoax.

In addition, in order to unravel what the allegedly public and harmful nature of counterfeit is made of, there is urgent need for ethnographic engagement in trying to understand the mechanisms of this affective and performative documentation work that is termed as predatory, *on their own terms*. Researchers looking for inspiration for such a mode of engagement and response may wish to look into recent work that blends art history with ethnography. For instance, Winnie Wong has examined Chinese Dafen "copyist" painters as they navigate a world where Western art is at once the gold standard and a commodity (2015). In Wong's work, Chinese hand-painted art products are observed ethnographically, and creators taken seriously for their craft and what they say about it, without moral judgment and without the filters of highbrow conceptual artists who unwillingly end up exoticizing them. What we find out through this engagement is that the privileged categories of originality, uniqueness, and authenticity are far more contingently constructed than we may think. In turn, the work that we associate with "fake art," that is, of manually copying, repetitively, and for pay, has in fact a lot in common with global contemporary art production in general. I can only surmise here, but given the current conditions of competitive, globalized science, it is not impossible that the actual activities of counterfeiters have more in common with those of "real scientists" that we can imagine.

Voices from the Global South also have to be included in policy research and policy-making debates on academic misconduct. Terms and themes engaging directly the Global South are almost completely absent from the conversation, including this book. Exceptions include Sarah de Rijcke and Tereza Stöckelová's contribution to the present volume, as they pointedly refer to the divide between the "international" West or North on the one hand and a "parochial" East or South in academic and publishing markets (de Rijcke and Stöckelová, this volume, chapter 7). When we think about issues such as public interest/public harm in research integrity, we have to reflect carefully about this divide and its distributive justice dimension. Science has a long history of translating "Third World people and their interests into research data within Western capitalist paradigms" (Escobar, 2011). Yet transnational and postcolonial critiques have not yet managed to infuse the organization, structures, and principles of research and publishing (Fletcher, 2015).

The processes of standardizing and measuring the forms science can take, and the privilege that comes attached to these, cannot be separated from the issue of counterfeit of the brands in science. Recognizing this fact highlights connections between normative good science and an ecosystem of scholarly publications that asserts privilege and exclusion. These connections are spelled out in other chapters of the present book (see de Rijke and Stöckelová).

Returning to Bohannon's sting, we see that the way it unfolded shows that within the scientific milieu, there seems to be a division between proper work of deception and improper work of deception. Bohannon and other authors of scientific media hoaxes, like Alan Sokal or more recently James A. Lindsay, Peter Boghossian, and Helen Pluckrose for instance (see also Lippman, this volume, chapter 21),[2] are perceived by many as being upright, brave deceivers who debunk and offer social criticism of sort, whilst also protecting the public. Interestingly it is assumed that Bohannon himself did not act fraudulently. We can ask at what point does the unmasking work of the denunciators who set traps to catch the improper deceivers become fraud itself? To answer this, it is worth paying attention to the format of the hoax as a strategy to expose fraud, as opposed to being an instance of fraud in itself. Journalist Curtis MacDougall's classic work *Hoaxes* defines the hoax as "deliberately fabricated falsehood made to masquerade as the truth,"[3] and philology attributes its origin to *hocus*, "to cheat."[4] The hoax can only work as a hoax if its author decides at the appropriate point in time to self-disclose and let others in. This temporality is critical, and the author of the hoax needs to maneuver it carefully, for the hoax would not work if it were its victims or a third party who would discover the plot. In cases where someone other than the author would reveal the hoax, its author could be considered as having committed deception, or fraud, just the same. Rhetorically speaking, there is no categorical demarcation between hoax and outright fraud, argues Lynda Walsh (2006). Whilst the hoaxer may be motivated by the desire to enact social criticism rather than rip people from their money or status, the authors of both hoax and fraud derive professional, reputational, or financial benefits, and inflict damage on their victims: wasting their time, causing reputational harm by depicting them as gullible, unprofessional, or vain (Walsh, 2006). Rereading the famous Sokal hoax with the tale of the emperor's new clothes, Walsh notes that the author of the hoax, Alan Sokal, demonstrated a desire to be seen as the canniest character, like the tailors. If the duped editors and peer reviewers of *Social Text* are cast as the emperor, whose vanity prevented candid admission of not understanding the article Sokal submitted, in this saga, Sokal self-posed as the clever and brave trickster who can tell us all how things really stand.

Looking at the "Authentic"

My second point is very simple and takes its cue from the previous chapter by Finn Brunton: in addition to creating welcomed opportunities for counterfeiters, the practice of counterfeiting also benefits those who are copied. Counterfeiting solidifies the "template" of elite science and keeps it intact; in other words, by reinforcing the value and prestige of the model, it often is "the sincerest form of flattery" (Mazzarella, 2015).[5] In our context, counterfeiters possibly contribute to sustain the structures of mainstream science by keeping them intact and off the radar whilst our scrutiny targets the counterfeiters. Further, through distinguishing themselves from the counterfeiter, the counterfeited—the elite journal, conference, or organization—accumulates further symbolic capital, as Adrian Johns has pointed out in his study of piracy (Johns, 2010). Therefore, the big challenge for the counterfeited is not quite identity theft as much as recuperating all that otherness, that externality associated with the counterfeiter, and using it tactically.

For watchdogs, including regulators and ethicists, the target remains the pirate, the predator, or the parasite. Whilst concerned with hunting misconduct, watchdogs pay less attention to the systemic features of the mainstream science on which the counterfeit models itself. Counterfeiters (and hoaxes, for that matter) often invite sustained scrutiny into the details and histories of relations and of hidden maneuvers. In the art world as much as in the scientific world, the work of copying is almost always condemned because it is not creative, not transformative or innovative, not critical, but a mere reiteration (Wong, 2013; Hayden, 2010). It is either feared as harmful or dismissed as useless. It is often mocked precisely because of the *modesty* inherent in this form of engagement. It is a "bad copy."[6] These get foregrounded when one examines counterfeiting, but relations are not dissected symmetrically when it comes to mainstream science.[7] Some of these relations are made explicit in Sergio Sismondo's contribution to the present volume (chapter 9; see Aldersey-Williams, 2005; Sismondo, 2009).

Let me illustrate further with an example from the Committee on Publication Ethics (COPE) Forum where (anonymized) allegations of breach of publication ethics get aired and debated amongst journal editors. Participants often state that they face a dilemma and thus have to choose between two potential goals: either solving disputes between authors/editors or maintaining the integrity of the "research record." In the former case, questions are asked about research funding and institutional arrangements. Uncovering the relations and processes that occur before and behind the publication of the paper in question is thought to be critical. "'Publication ethics' does not come out of nowhere," a COPE governing member says, acknowledging explicitly that in order to "do" publication ethics by way of helping resolve a dispute between authors, one has to take stock of a composite of different persons and roles as well as institutions, some legitimated and some less. For instance, different forms of authors, including guest or honorary, ghost, external consultant, medical writer, and student, operate within structures where the line between pure academic work and market-driven research may no longer exist (Rabinow, 1996), but where hierarchical lines of authority between established professors, domestic and international, English-speaking and non–English-speaking PhD students, and early-career and experienced scholars still hold sway. This unpacking takes place when the COPE members discuss authorship dispute. However, when discussions deal with cases of alleged falsification or fabrication of data, participants tend to construct the research record as a self-contained object, and emphasize the need for maintaining its inherent integrity. In these cases, the "research record" is made into an object, detached from, but possibly threatened by, supposedly external personal relations or histories. If the research record is threatened by misconduct, it can, in turn, be restored as a standalone object. My research shows that this process of restoration reifies the research record, isolates it from human relationships (between authors, or between authors and editors) and, in turn, makes these relations recede in the background (Jacob, 2019).

So publication ethics watchdog organizations like COPE struggle to get the full picture when it comes to counterfeit and to mainstream science, but like most other players in the milieu, they are also stuck, albeit unwillingly, in the loop that links together authentic and counterfeit.

Ethics Can Be Counterfeited Too

The publication ethics and research integrity movement demands authenticity within a system that has conflicting demands over value. In many ways and as this book makes explicit, the ecology of science, by demanding both quality and high quantity from players (academic authors), drives the demand for fake. So watchdogs are extremely busy. Reflecting upon the tension between fake and authentic, and the regulatory activities this tension entails, leads me to another analogy from outside of scientific publishing: the consumer movement's response to the market in faked goods in China.

Anthropologist Susanne Brandtstädter has researched a citizen-led movement that acts as a sort of watchdog against counterfeit products in China. Brandtstädter's intriguing work shows that "value" as quality and authenticity is itself also a currency that can be accumulated, invested, and distributed (2009). The contradiction of demands—*demands for fake* and *demands for true value by the consumers* in China—means that local stall owners now aim to cater to both by producing a fake Gucci bag that looks real, but importantly, that also comes with a fake certificate of authenticity (2009). Similarly, certifications and signatures are a big part of the added value of Dafen hand-painted art products (Wong, 2013). But why is this observation interesting for our thinking about publication ethics watchdogs? It means watchdogs' brands are also at the risk of being counterfeited. Certifications of "ethics," "authenticity," and "integrity" have become templates that can become vulnerable as such. This is not unique to publishing since "ethic" is a fruitful template to replicate, in many areas (ethical certification is used for organic food and fair trade, for example).

COPE, for instance, aims to provide an example of good practice and professionalism within publishing. The example it provides is activated through material objects like its flowcharts, newsletters, and, of course, its logo. Its logo is a mark, a kind of certification with its own aura, which itself can be counterfeited. To preserve the authentic nature of its name and logo, COPE recently transformed its logo into multiple personalized logos that each contains a unique number, now available for download by their genuine, fee-paying registered members—out of painstaking concern for preserving a brand of "authenticity" that is vulnerable.

Brand names allegedly "concern the need to provide information to consumers/readers/citizens efficiently about the unobservable qualities of the product that is being sold" in order to assist decision making (Copeman and Das, 2015). But of course they do more than that: brand names try to create markets for products (Mazzarella, 2015). Any product. With this unavoidably comes the mimetic ability to copy the name and use it for an inferior product, or to take a similar-sounding name and thus to steal a part of the name and the market share (Copeman and Das, 2015). What is most crucial to recognize here is the inevitability rather than exceptionality of the eruption of counterfeit scientific journals. It is unhelpful to see the "make believe" as anomalous.

Notes

Acknowledgments: I wish to thank the editors Mario Biagioli and Alexandra Lippman for inviting me to contribute and for their helpful feedback.

1. John Bohannon explains: "My hope was that using developing world authors and institutions would arouse less suspicion if a curious editor were to find nothing about them on the internet."

2. The Sokal affair refers to a scientific publishing hoax perpetrated in 1996 by professor of mathematics Alan Sokal. Sokal submitted a nonsensical article entitle "Transgressing the Boundaries: Towards an Hermeneutics of Quantum Gravity" to the academic journal of cultural studies *Social Text*. With this submission, he wanted to conduct an experiment to test the journal editors' rigor and to see whether the article could get published. The article did not undergo peer review but was accepted by the editors and published in *Social Text*'s special issue on the science wars. On the day of publication, Sokal wrote a piece in *Linga Franca* disclosing the hoax. The hoax triggered many debates within and beyond academia on publishing ethics, postmodernism, and rigor in the humanities. In 2018 Lindsay, Boghossian and Pluckrose conducted a hoax on what they call grievance studies scholarship and peer-review process. The hoax has been called Sokal Squared in reference to Sokal's hoax.

3. Curtis MacDougall, *Hoaxes*, Dover, 1958.

4. Robert Nares, 1822, "A Glossary; or, Collection of Words Which Have Been Thought to Require Illustration in the Works of English Authors," London: Robert Triphook.

5. The proliferation of counterfeits may divert some customers away from the authorized product while at the same time heightening the prestige of "the real thing" (Mazzarella, 2015).

6. Cori Hayden, 2010, "The Proper Copy," *Journal of Cultural Economy* 3(1):85–102.

7. See Bruno Latour, 1991, *Nous n'avons jamais été modernes*, Paris: La Découverte.

References

Bohannon, John. 2013. "Who's Afraid of Peer-Review." *Science* 342(6154):60–65.

Brandtstädter, Susanne. 2009. "Fakes: Fraud, Value-Anxiety, and the Politics of Sincerity," in Karen Sykes, ed, *Ethnographies of Moral Reasoning: Living Paradoxes of a Global Age*. Basingstoke: Palgrave, 139–160.

Copeman, Jacob, and Veena Das. 2015. "Introduction. On Names in South Asia: Iteration, (Im)propriety and Dissimulation." *South Asia Multidisciplinary*

Academic Journal [Online], December 2015, accessed May 10, 2016. http://samaj.revues.org/4063.

Craciun, Elena M. 2012. “Rethinking Fakes, Authenticating Selves.” *The Journal of the Royal Anthropological Institute* 18(4):846–863.

Escobar, Arturo. 2011. *Encountering Development: The Making and Unmaking of the Third World*. Princeton, NJ: Princeton University Press.

Fletcher, Ruth. 2015. “Flak: Mixing Feminism, Legality and Knowledge.” *Feminist Legal Studies* 23(3):241–252.

Hayden, Cori. 2010. “The Proper Copy.” *Journal of Cultural Economy* 3(1): 85–102.

Jacob, Marie-Andrée. 2015. “Misconduct Hunting: Research Integrity via Law, Science, and Technology,” in E. Cloatre and M. Pickergsill, eds, *Knowledge, Technology, and Law*. Abingdon: Routledge, 137–152.

Jacob, Marie-Andrée. 2016. “On the Scope and Typology of ‘Research Misconduct’: The Gaze of the General Medical Council, 1990–2015.” *Medical Law Review 24(4):497–517.*

Jacob, Marie-Andrée. 2017. “The Strikethrough: An Approach to Regulatory Writing and Professional Discipline.” *Legal Studies* doi:10.1111/lest.12142.

Jacob, Marie-Andrée. 2019. “Under Repair: A Publication Ethics and Research Record in the Making.” *Social Studies of Science* doi: 10.1177/0306312718824663.

Jingfeng, Xia, Jennifer L. Harmon, Kevin G. Connolly, Ryan M. Donnelly, Mary R. Anderson, and Heather A. Howard. 2015. “Who Publishes in ‘Predatory’ Journals?” *Journal of the Association for Information Science and Technology* 66(7):1406–1417.

Johns, Adrian. 2010. *Piracy: The Intellectual Property Wars from Gutenberg to Gates*. Chicago: University of Chicago Press.

Latour, Bruno. 1991. *Nous n’avons jamais été modernes*. Paris: La Découverte.

Mazzarella, William. 2015. “On the Im/Propriety of Brand Names.” *South Asia Multidisciplinary Academic Journal* 12 [Online], September 21, 2015. http://samaj.revues.org/3986.

Navarro-Yashin, Yael. 2012. *The Make-Believe Space: Affective Geography in a Postwar Polity.* Durham, NC: Duke University Press.

Rabinow, Paul. 1996. “American Moderns: On Science and Scientists,” in *Essays on the Anthropology of Reason*. Princeton, NJ: Princeton University Press, 162–187.

Siegel, James. 1998. *A New Criminal Type in Jakarta*. Durham, NC: Duke University Press.

Walsh, Lynda. 2006. *Sins Against Science: The Scientific Media Hoaxes of Poe, Twain and Others*. Albany, NY: SUNY Press.

Wong, Winnie. 2014. *Van Gogh on Demand. China and the Ready-Made*. Chicago: University of Chicago Press.

20

Fake Archives: The Search for Openness in Scholarly Communication Platforms

Alessandro Delfanti

Doppelgängers exist in science too. While counterfeit or reappropriated brands have long been regular features of the commercial world, scholarship is increasingly dealing with a similar phenomenon. Appropriating a successful brand is indeed a way to tap into the value the original brand creates and the communities it fosters. This does not need to be illegal or deceptive though. viXra.org is a preprint repository that mimics the design, logo, structure, and functioning of arXiv.org, the open-access website that collects articles from physics, mathematics, and other quantitative sciences before or regardless of their submission and publication in a peer-reviewed journal. Launched in 2009 as an answer to the role of arXiv as the dominant platform for scholarly publishing in some areas, viXra is an ironic copycat version of the "official" website, of which it spells the name backwards (Brumfiel, 2009). At the same time though, one cannot help but notice that viXra contains thousands of articles. It has in fact grown to become an alternative platform for scholarly communication. While it would be easy to discard it as a container for "crackpot" and irrelevant science such as cold fusion or unorthodox astrophysical theories, it hardly represents a form of misconduct. Sure, viXra engages in spam-like practices (Brunton, this volume, chapter 18). For example, users who misspell arxiv.org's URL and type "rxiv.org" instead will land on a website that mirrors viXra's and presents its content.[1] But what is more important is that viXra could help shed light on how current forms of digital scholarly publishing run counter to rhetorics of openness, and how practices of brand appropriation and mimicry can allow criticism to be embodied by concrete, if ironic, alternatives.

Like the made-up scientist Ike Antkare (this volume, chapter 14), viXra seems indeed to aim at highlighting some of the critical issues at stake in a media ecology in which digital platforms for publishing and valorizing scholarly content are assuming an increasingly central role.

These platforms may be open for any reader to access their content free of charge, but can also entrench other forms of power. Indeed, viXra's website bears the motto "open e-print archive" as a remembrance of its critique of the "official" repository's publication standards. In many cases, when we think about ways of gaming metrics of scholarly output we refer to peer-reviewed journal articles and citations, which tend to be seen as the main factor underpinning academic rankings and evaluations. Yet there are other forms of publishing that have come to represent crucial places through which academic credit is built. Among these, online repositories and social media such as Academia.edu or Social Science Research Network (SSRN) are emerging as key spaces of both knowledge circulation and credit allocation. These platforms come with two relevant characteristics. First, they tend to become inter- or intradisciplinary powerhouses and thus obstruct the emergence of competing actors such as new platforms but also, for example, predatory or irrelevant journals. Second, they come with their own sets of detailed metrics, such as download counts, internal citations, popularity, and other rating systems, allowing for individual forms of metrics microcontrol. This has been dubbed "gamification of research" (Wagman, 2016).

In physics and mathematics, arXiv.org presents both characteristics and thus represents a unique bottleneck. Founded by particle physicists in 1991 and now run by Cornell University, arXiv is the hegemonic space of circulation for scholarly articles in a number of scientific disciplines. After its launch in particle physics in 1991, arXiv has quickly expanded to cover a number of subfields. In particle physics, for example, it quickly plateaued to include more than ninety percent of all articles published in the field (Gunnarsdottir, 2005). In 2014, it passed the mark of one million deposited papers across the disciplines it serves. This repository grew out of traditional epistolary exchange practices that predate digital communication technologies and have been institutionalized in a "preprint culture" since the end of World War II, especially in physics. In some of the disciplines covered by arXiv, such as particle physics, publications in peer-reviewed journals are recognized for prestige and recognition outside the community, but arXiv provides quick and broad recognition within it. One could say that in the eyes of the community, a physicist does not exist if their work does not appear on arXiv. Physicists and mathematicians simply refer to it as "the archive." The website publishes preprint versions of scholarly articles and at the same time provides metadata for platforms that provide metrics of impact: for example, Google Scholar or inSPIRE, a particle physics service that uses arXiv data to provide author pages and metrics such as citation counts. Citations allow inSPIRE to rank articles on a ladder that starts with “renowned,” moving to categories such as “famous” or “well known,” and ending with “unknown” papers.

Scholarly Brands

The backward-spelled viXra is the evil twin that appropriates arXiv’s name and image. Theories about brands have been developed for marketable products that are constructed, approached, accessed, and used very differently from a scholarly journal or publishing platform. Yet focusing on viXra as a copycat version of the brand arXiv might help understand their relationship. Brands are incredibly powerful. They make up a relevant chunk of many companies’ value. Also, far from being mere pieces of design, brands are underpinned by socially meaningful activities such as customer activity on social media or informal product re-elaboration, just to mention a couple of examples. Since their value is at the very least partially the product of customer activities, brands need to be flexible enough to modulate, incorporate, and valorize such activities (Arvidsson, 2006). Thus brands are able to incorporate innovation while at the same time being subject to hijacking. Brands are thus at the same time extremely powerful and extremely weak.

For example, brands can be re-elaborated and thus made more authentic by customers—think of the sticker that makes the illuminated apple on a Mac laptop look like as if it is being held by Snow White’s witch. These activities have a direct effect on a brand’s value. In other cases, this very flexibility can expose brands to appropriation by actors the company cannot control. Counterfeit brands can come to existence when an original product with a remarkable brand value worth copying already exists on the market. Its characteristics are copied into another product, which is at least partially indistinguishable from the original, and is sold at a lower price. Oftentimes consumers are well aware of the difference between the two products. Counterfeit products would be a clear example of partial appropriation of a brand’s value by external actors. You can buy copycat Nike shoes carrying a perfect “swoosh” logo and none of that money will go to Nike. Counterfeit journal websites routinely collect author fees from scholars deceived by doppelgänger versions of recognized journals. In many cases, these hijacked journals use slightly different names and graphics than the originals, masquerading as “legitimate” journals such as *Archives des Sciences* or *Wulfenia* (Butler, 2013).[2]

Scholarly journals continuously engage in battles over the identification and denunciation of these copycat scam operations. But this is not the main aspect of brand repurposing in viXra's case.

A different case of brand appropriation is represented by brands that purposely tweak, or subvert, the original brand. Through practices of *subvertising* (a portmanteau of *subvert* and *advertising*), political activists transform a logo and use it against the original, for political purposes. Oftentimes this means exposing issues of concern such as labor conditions or environmental problems caused by the company (Klein, 1999). Think of the golden arches of McDonald's logo reading "McDiabetes." Subverted brands are not only tools to be used in the public sphere to criticize something, they can also be used to organize new publics and thus create alternative spaces or entities that are based upon arrangements that contradict a brand's logics. This kind of reappropriation functions according to a sort of "the medium is the message" logic, whereas the very existence of the subverted brand may be more important than the content it carries. This could be the case for viXra, which is both a critique and an alternative to the "official" archive and is not meant to deceive its users. As stated on viXra.org: "The similarity of web design is a form of parody to highlight the endorsement and moderation policies of arXiv.org which we believe are a hindrance to scientific progress. We reverse the name and colours as a symbol of our opposing policies and to ensure that there is no confusion between the sites."[3]

The hindrance mentioned by viXra is created by the ways in which arXiv.org is made available to some scholars while fencing off others. What is at stake here is the meaning and role of "openness" in contemporary scholarly communication.

arXiv vs. viXra

Since arXiv.org is so central to the physics community, people have started asking how it shapes publishing and evaluation practices within the field. Indeed, arXiv is the place where credit is attributed and community boundaries are continuously created (or perhaps we should say enforced), especially within some subfields, such as particle physics, which put less emphasis on journal publications (Delfanti, 2016). Its backbones are its technological infrastructure and its moderators, chosen within the disciplinary communities it serves. Moderators control the quality of the submitted articles and thus guarantee the scientific relevance of arXiv's content. The archive is considered one of the flagship infrastructures of the open-access movement. Indeed, arXiv.org is free of charge and open access. Anyone with a computer connected to the internet can access and download its content without any restrictions such as paywalls or registration. Nevertheless, what is not fully open is arXiv's submission and publishing system. ArXiv is open for readers, but not for authors. Unlike repositories such as Academia.edu or SSRN, which allow scholars to publish preprints or peer-reviewed articles and do not filter content, arXiv is not a self-publishing platform. It checks submissions for quality, enforcing rules that would be deemed inadequate for scholarly journals.

There are indeed three filters, all geared toward making sure that the content amounts to "physics" or "mathematics" and is original. First, in order to publish on arXiv, one needs to set up an account, which can only be done through a recognized institutional email account, such as utoronto.ca, or by receiving an endorsement from an active arXiv user. Second, a machine learning software performs automated textual analysis, scanning articles for plagiarism or technical issues, and sorting out papers that might be uninteresting. Third, articles "flagged" by this system are checked by human moderators. As a result of this process, articles can be outright rejected, although this seems to be rare; they can be redirected to categories such as PH-GEN (general physics), a "ghetto" subcategory that tends to include papers that are authored by physicists that are recognized as members of their disciplinary community, but do not live up to the standards of publication; or they can be accepted for the required specific category, such as HEP-TH (theoretical high-energy physics) (Reyes-Galindo, 2016). Arguably, arXiv does a great job of including recognized members of the specific research communities it hosts, and institutionally sanctioned physicists and mathematicians tend to agree that it is a fantastic platform for fast, reliable, and relevant communication. At the same time, criticism may be rare, but it strikes to the core of arXiv's functioning and role. Over time, individual members of the research community have criticized arXiv for its lack of transparency or for blacklisting scholars (a practice arXiv denies; Merali, 2016). The website *arXiv Freedom* lists cases of independent scientists who accuse arXiv of "abuse" for rejecting articles that are deemed uninteresting or fringe.

Out of the frustration with arXiv's dominant role and in response to its perceived abuses, in 2009, physicist Philip Gibbs created viXra. This is a doppelgänger of the "official" archive, a clone website that is identical to arXiv, including many of the same categories, such as astrophysics or condensed matter, the same logo (spelled backwards and in a different color), and the same organization and presentation of published papers.

While stressing that viXra is a parody, the founders aimed at building an alternative archive that could be open to “the *whole* scientific community” (emphasis added). The founders of viXra believe that, in contemporary scholarly communication, openness should not stop short of allowing all researchers to place their ideas in public view for scrutiny. Indeed, descriptions on the viXra.org website state it is “an experiment to see what kind of scientific work is being excluded by the arXiv. But most of all it is a serious and permanent e-print archive for scientific work. Unlike arXiv.org it is truly open to scientists from all walks of life.”[4]

While viXra’s success is based on its explicit mimicry of arXiv, its ability to create a community should not be overlooked. We know from marketing studies that preference for a counterfeit brand is greater when the brand attitude serves a social function. This means that alternative brands are accepted and embraced when they help people construct and maintain social bonds, and thus create new collectives. Counterfeit brands help people maintain relationships. For example, consumers are motivated to consume a copycat product to gain approval in social situations rather than communicate their central beliefs, attitudes, and values to others (Wilcox, 2009). The collective organized around viXra may indeed sound like a strange one. But while viXra’s main feature may be the way it uses parody to strike a critique at the core of the functioning of digital preprint archives, the fact that it is not just a boutade or media performance is crucial. As of June 2017, viXra had gathered more than eighteen thousand articles, the content of which is composed by lots of unorthodox or “crackpot” science. But this may be beyond the point: although most authors seem to be nonacademically trained, viXra also contains articles that have been published by institutionally recognized, peer-reviewed scholarly journals, as well as articles that are cross-posted on arXiv (Kelk and Devine, 2012). Indeed, aiming at gathering science “from all walks of life” leads viXra to attract a very diverse mix of content, arguably with a majority of ideas that would not be considered appropriate within institutionally sanctioned physics. In a sense, the repository recognizes that this is not what makes viXra relevant. Indeed, it explicitly states it is not interested in competing with scholarly journals or becoming a space for credit attribution and reputation building: “Acceptance into viXra does not constitute a publication of research in the academic sense since no quality review takes place.…ViXra does not aim to improve its reputation by filtering by quality. Our aim is to cultivate a reputation for openness by supporting free speech principles in science.”[5]

But does viXra not constitute a publication? Arguably that is exactly what the "official" archive does: while it can not claim to produce institutionally recognized publications, it uses community-based forms of recognition to attribute credit, at least within a specific community, to scholarly objects that have not been subject to formal expert peer review. Openness is a concern because the lack thereof in arXiv's screening process is meant exactly to protect this system of community-based recognition. On the flip side of the coin, scholars working in institutional settings seem to be discouraged from publishing on viXra. This could indeed affect their credibility. But why? After all, it has become increasingly normal for scholars to publish their work on institutional repositories or academic social media such as ResearchGate or Academia.edu. Credibility issues are still present within some disciplines, but publication on other platforms does not seem to be as heavily moralized as it is with viXra. This may be because publishing on viXra is not a matter of irony but rather a way of joining another community that does recognize those papers as publications. Also, viXra struggles with making its content accepted by services that could use its metadata to provide metrics of impact. Neither Google Scholar nor inSPIRE provide information on viXra papers. Instead, viXra provides download statistics for individual articles, thus forging its own metrics system, albeit arguably an outcast one.

Conclusions

If the "official" arXiv is a central space for the attribution of individual credit and the emergence of metrics of impact such as citation counts, the alternative repository viXra tries to criticize and at the same time emulate this role. Poking fun at arXiv is common in physics. The *snarXiv* is a "random high-energy theory paper generator" that mimics arXiv while producing articles generated by randomly aggregating particle physics lingo. Its *arXiv vs. snarXiv* page is a web-based test that asks you to pick the "real" article when confronted with a title from arXiv and one randomly generated, for example, "An Entropic Resolution of the Confinement Problem Magnetic" vs. "Monopole Is Photon."[6] Yet while these forms of brand appropriation amount to jokes based on a shared subculture, viXra manages to highlight how the use of preprint archives should be analyzed in the light of their role as dominant keepers of the boundaries of a scholarly field. ArXiv's systematic exclusion of nonrecognized scholarship is based on principled decisions that have to do with system efficiency and noise reduction, as well as with the maintaining of closed community boundaries. Also, it has created its own measurements of impact, thus shaping publishing and evaluation practices.

Obviously, the history of digital cultures has shown us that projects that seemed weird and marginal at their birth, like Linux or Wikipedia, ended up reshaping their fields. They did so precisely by opening up to broad constituencies that were not part of incumbent institutions—think IBM or Encyclopedia Britannica. This analogy would be amiss though. Sure enough, the alternative repository viXra shows that the rhetoric that portrays digitally enabled communication as providing universal open access to the scholarly communication system may be misplaced. It also provides an alternative for noninstitutionally sanctioned scholars who need a preprint publishing system in order to become visible and gain access to a community. So on the one hand, viXra could be seen as an attempt—funny and goofy perhaps—at using a copycat platform to contribute to shaping the evolution of the original one (Jacob, this volume, chapter 19) or of scholarly publishing more in general. One could argue instead that viXra is just a weird space where bogus and crackpot physics and mathematics find their way through a fake system of publication that could not even aspire to misconduct. Yet its most important (or sole?) accomplishment may be the critique it embodies by parodying arXiv. As a subverted brand, viXra highlights the increasingly crucial role played by non–peer-reviewed, preprint-based spaces, with their own gatekeeping rules and metrics of impact. As these platforms are quickly establishing themselves as dominant actors within some disciplines (for example, SSRN for law, Academia.edu for many areas of the social sciences), critical practices such as brand appropriation may help expose the limits of these publishing venues as well as the new forms of power that they are entrenching. Perhaps viXra will help us imagine how to build the real alternatives we need so badly.

Notes

1. http://www.rxiv.org/ (last accessed June 13, 2017).
2. For an updated list of hijacked journals, see http://beallslist.weebly.com/hijacked-journals.html (last accessed June 13, 2017).
3. http://www.vixra.org/disclaimer/ (last accessed June 13, 2017).
4. http://www.vixra.org/why (last accessed June 13, 2017).
5. http://www.vixra.org/info (last accessed June 13, 2017).
6. http://www.snarxiv.org (last accessed June 13, 2017).

References

Arvidsson, Adam. 2006. *Brands: Meaning and Value in Media Culture*. New York: Routledge.

Brumfiel, Geoffrey. 2009. “What’s arXiv Spelled Backwards? A New Place to Publish.” *Nature News*, July 16, online.

Butler, Declan. 2013. “Sham Journals Scam Authors.” *Nature* 495(7442):421.

Delfanti, Alessandro. 2016. “Beams of Particles and Papers: How Digital Preprint Archives Shape Authorship and Credit.” *Social Studies of Science* 46(4):629–645.

Gunnarsdottir, Kristrun. 2005. “Scientific Journal Publications On the Role of Electronic Preprint Exchange in the Distribution of Scientific Literature.” *Social Studies of Science* 35(4):549–579.

Kelk, David, and David Devine. 2012. “A Scienceographic Comparison of Physics Papers from the arXiv and viXra Archives.” arXiv preprint, arXiv:1211.1036.

Klein, Naomi. 1999. *No Logo: Taking Aim at the Brand Bullies*. Toronto: Knopf Canada.

Merali, Zeeya. 2016. “ArXiv Rejections Lead to Spat Over Screening Process.” *Nature News*, January 29, online.

Reyes-Galindo, Luis. 2016. “Automating the Horae: Boundary-work in the Age of Computers.” *Social Studies of Science* 46(4):586–606.

Wagman, Shawna. 2016. “Some Academics Remain Skeptical of Academia.edu.” *University Affairs*, April 12, online.

Wilcox, Keith, Hyeong Min Kim, and Sankar Sen. 2009. “Why Do Consumers Buy Counterfeit Luxury Brands?” *Journal of Marketing Research* 46(2):247–259.

21

Humor, Hoaxes, and Software in the Search for Academic Misconduct

Alexandra Lippman

The Office of Research Integrity is hardly renowned for its sense of humor. By contrast, a new generation of independent watchdogs and bloggers interested in academic misconduct employ jokes, pranks, witty pseudonyms, and humorous hoaxes as a part of their critique and as tools of investigation. The new watchdogs represent a significant shift from top-down, bureaucratic, institutionalized detection of academic misconduct toward collaborative discussion, detection, and dissemination. Not only is this work often done for free, but it also is often done with and through humor.[1]

Until recently, hoaxes within academia targeted authorities—highly regarded scholars, journals, or disciplines. In 1996, New York University physicist Alan Sokal wrote and submitted an article to *Social Text*. His goal was to test whether a top cultural studies journal in the United States would publish an article rife with nonsensical claims "if (a) it sounded good and (b) it flattered the editors' ideological preconceptions" (Sokal, 1996). But unlike Sokal's relatively straightforward hoax, which *Social Text* accepted and published, the bloggers and pranksters discussed in this chapter reveal shams through elaborate jokes and stings, crowd participation, and the creation of fictional personae. Furthermore, while hoaxes as a genre target reputable institutions and figures of authority, contemporary misconduct watchdogs' hoaxes and jokes take aim at fraudsters with a sense of humor.[2]

I argue that the detection, critique, and mocking of academic gaming and scholarship have taken a carnivalesque turn. I will discuss changes within the focus of critique by comparing two software-based, scholarly article generators released a decade apart: the Postmodernism Generator, created in 1996, and SCIgen, created in 2005. While the creator of the Postmodernism Generator playfully mocks renowned humanities scholars' jargon through producing Dadaist computer-generated papers, the creators of SCIgen aim lower and mock spammish[3] conferences by creating computer-generated computer science papers. Through mimicking academic articles, article generators raise questions about the style and content of academic writing, reading, and peer review. As we argue in our introduction to the volume, the increasing emphasis on scholarly metrics has led to the rise of a new type of unambitious academic misconduct, which focuses on gaming metrics or publishing in spammish journals to meet basic requirements for academic employment, retention, and promotion. Academic watchdogs and scholar-pranksters track this shift by humorously targeting fraudsters and so-called predatory journals instead of figures of authority.

Carnivalesque Watchdogs

Contemporary misconduct watchdogs—without institutional affiliation, unconnected to funding agencies—mark a transformation from top-down to bottom-up knowledge production around emerging forms of academic misconduct and metrics-driven gaming. I contend that their work tends toward the carnivalesque since it is marked by crowd participation, anonymity, and various uses of humor.[4]

Crowd-generated knowledge and rumors drive discussion and detection of misconduct. The website Retraction Watch asks the public for "any tips … about the nature of a retraction, expression of concern, or correction" (Oransky, this volume, chapter 10). The short-lived blog "Science Fraud" also used crowdsourcing and knowledge sharing about misconduct. Paul Brookes, a biochemist at the University of Rochester, ran the blog under the pseudonym "Frances de Triusce" (an anagram of "science fraudster") (Brookes, this volume, chapter 13). Numerous readers emailed him examples of suspected data irregularities and image manipulation. During the six months of its existence, Science Fraud discussed 275 articles (of which 16 were retracted and 47 issued corrections) and credited tipsters by their chosen pseudonym, thanking "*freddy fraudster*," "an astute reader," and "*Blotette* (again!)." In part, Science Fraud—which relied on anonymous discussion—was motivated by Brookes's frustration with the ORI's slow pace and particularly with their revelation of his identity to the mentor of a postdoctoral scholar, in whose work he had discovered image manipulation. The ORI's unmasking of Brookes's identity to the mentor of the scholar accused of fraud pushed Brookes to pursue a mode of critique that would preserve anonymity of participants (including himself).

Watchdogs' anonymity and pseudonimity can create a carnivalesque atmosphere. Anonymity creates a freer space for watchdogs' discussion of and jokes about suspected misconduct. This turn toward disguise recalls Mikhail Bakhtin's discussion of the symbolic meanings of masks used during carnival: "The mask is related to transition, metamorphoses, the violation of natural boundaries, to mockery and familiar nicknames. It contains the playful element of life; it is based on a peculiar interrelation of reality and image" (1984). Pseudonyms, anonymity, and masked IP addresses recall the masks of carnival. Disguises facilitate a range of humorous responses from playfulness, to critique, and to mockery.

Sometimes—as is characteristic of carnival—the watchdogs poke fun at powerful authorities. Anonymity makes it possible for junior scholars to critique established academics without fear of repercussions. For instance, anonymous discussions on Retraction Watch and PubPeer of leading French plant biologist Olivier Voinnet's work revealed his manipulation of images and resulted in his being suspended for two years from the French National Centre for Scientific Research (Guaspare and Didier, this volume, chapter 12).

The online journal club PubPeer allows the public to comment on published papers either anonymously or with their names on their website.[5] Founded by Brandon Stell and brothers George and Richard Smith in 2012, PubPeer's stated goal "is to foster a scientific environment where robust, high-quality research is valued, while providing a forum to discuss the problems of unreproducible, misleading, misconceived, or fraudulent work."[6] The post-publication peer review allows for the in-depth discussion of articles.[7] While PubPeer allows for the public to comment anonymously, the tone of the discussion—unlike other watchdogs I will discuss—remains fairly serious.

The high seriousness associated with university committees and governmental institutions dedicated to ethics and research integrity contrasts with the carnivalesque attitude pervasive in crowdsourced misconduct detection. For instance, Science Fraud titled posts with puns of scientists' names. One blog post titled "Ezzat a paper in your pocket or are you just pleased to see me?" highlighted oncologist Sheeran Ezzat's "creative imagery."[8] Retraction Watch is equally fond of puns in the titles of their postings: "Warts and All: Derm Pub Retracts Plantar Paper after Author Cries Foul," or "Double trouble: Psych Journal Prints PTSD Paper Twice." Retraction Watch also posts a "leaderboard," ranking authors with the most papers retracted—a snide inversion of the impact factor.[9]

Choosing pseudonyms also offers an opportunity for humor and mockery. Computer scientist Cyril Labbé invented "Ike Antkare" (I Can't Care) as the fictional author of multiple SCIgen papers, and successfully skewed Google Scholar's ranking to make Ike Antkare rank above Albert Einstein (Antkare, this volume, chapter 14). Similarly, when Burkhard Morgenstern created fictional scholars to join spammish journals' editorial boards, he chose ridiculous photos, biographies, and names like Hoss Cartwright from the television series *Bonanza* (Morgenstern, this volume, chapter 15). Dr. Cartwright with his postdoctoral fellowship at "Cowboy University" and current position as "senior cattle manager" graced the editorial boards of various journals from the *International Journal of Agriculture Innovations and Research* to the *Journal of Primatology.*

Hailing from a variety of backgrounds and professional trajectories, the new generation of watchdogs democratizes participation in the detection and discussion of misconduct and blurs the line between "policemen" and "pranksters." This anti-hierarchical tendency also recalls carnival, which, as Bakhtin asserts, "does not acknowledge any distinction between actors and spectators.… Carnival is not a spectacle seen by the people; they live in it, and everyone participates because its very idea embraces all the people" (Bakhtin, 1984). Carnivalesque critique—facilitated through anonymity and disguises—permits satire, pranks, and derision from the crowd.

Mocking Authority: The Postmodernism Generator

In 1996, a few months before Sokal's hoax, Andrew C. Bulhak released the Postmodernism Generator and the underlying Dada Engine software. Richard Dawkins calls the generator "a literally infinite source of randomly generated, syntactically correct nonsense, distinguishable from the real thing only in being more fun to read" (1998). Bulhak claims inspiration from Douglas Hofstadter's best-selling (and Pulitzer Prize–winning) *Gödel, Escher, Bach: An Eternal Golden Braid* (1979), in particular the author's ideas about recursion. Hofstadter demonstrates a method for producing grammatically correct—but nonsensical—English texts and illustrates it "with an example, a selection of fragments of text, ten of which were generated using a computer program and three which were taken from a journal titled *Art-Language*[,] and a challenge to the reader to identify which ones were generated artificially" (Bulhak, 1996).

The Postmodernism Generator mocks famous scholars of postmodernism, cultural theory, and literary criticism. Bulhak chose these genres because he thought it "easy to convincingly generate meaningless and yet realistic travesties of works in it" (Bulhak, 1996). Bulhak imagined that "automated travesties of papers in, say, mathematics or physics, would be less successful because of the scientific rigor of these fields" (1996). By contrast, he also discusses how the Dada Engine can be modified to simulate the "ranting of a paranoid schizophrenic street preacher, or perhaps a USENET ranter" or to generate "eccentric pseudoscientific/religious pamphlets" (1996). By titling his technical report "On the Simulation of Postmodernism and Mental Debility using Recursive Transition Networks," Bulhak draws parallels between the two. He notes that the patterns of "abnormal modes of human communication—such as a restricted specialized field of discourse…—are easier to replicate than normal communication" (1996).

The version of the generator that I accessed online has delivered 14,440,299 essays since it went live in 2000. When I opened the website, Elsewhere.org/pomo, I generated an essay titled, "The Collapse of Culture: Baudrillardist simulacra and constructivism," co-authored by two nonexistent professors at the University of Illinois and U Mass-Amherst. This essay included lines such as the following: "'Sexual identity is part of the economy of truth,' says Sontag. If predialectic textual theory holds, we have to choose between capitalist theory and neocultural narrative. Thus, Lyotard uses the term 'Baudrillardist simulacra' to denote a self-sufficient totality."

Theorist name-dropping, scholars' names transformed into adjectives—for example, Foucaultian, Debordist—citations, footnotes, sections, and bibliographies all simulate specialized writing, while the generation of the essay through a software program mocks the genres. Through its travesties of papers, which reference major figures of cultural studies and literary theory, the Postmodernism Generator provokes laughter at authorities and fits traditional carnivalesque critique, which aims high rather than low.

Spamming and SCIgen

In 2005, three graduate students at MIT's Computer Science and Artificial Intelligence Laboratory decided that they had had enough of spammish calls for papers flooding their inboxes. In the tradition of many MIT students, they set up a clever prank. They created SCIgen, software that generates computer science papers, which they could then submit to spammish conferences. The students, Dan Aguayo, Max Krohn, and Jeremy Stribling, worked on the project for a couple of weeks. The capaciously named World Multiconference on Systemics, Cybernetics, and Informatics (WMSCI, 2005) accepted one of their SCIgen contributions, "Rooter: A Methodology for the Typical Unification of Access Points and Redundancy" as a "nonreviewed" paper for their conference.

After the hoax was revealed, WMSCI withdrew their paper. Like persistent spammers, however, the graduate students did not give up. They raised money to attend and set up a parallel session at the conference hotel in Orlando entitled "Methodologies, Theory, and Information" to prank the conference. Although WMSCI attempted to forbid them from using its name on their fliers and posters advertising their session, the determined grad students persisted. Sporting fake mustaches, wigs, clip-on ties, and white lab coats, Aguayo, Krohn, and Stribling presented from SCIgen slides, which they had not read beforehand.[10] Trying to keep a straight face, they discussed slides with conclusions like:

- Scherzo will address many of the obstacles faced by software engineering
- Prevents the World Wide Web
- We verified that redundancy and e-business can interfere to achieve this aim
- Our application represents a profound advancement to theory

A couple bewildered people sat through some of the PowerPoint karaoke. By wearing costumes, using silly pseudonyms, and reading ridiculous SCIgen presentations, the three MIT students pranked the conference.

SCIgen's success and popularity derives in part from its being very funny. Its creators wrote humor into the code. Aguayo, Krohn, and Stribling explicitly avoided more sophisticated, technically challenging approaches such as Markov chains because, although they could produce proper syntax, the material could turn out to be dry and boring. In an "Ask Me Anything" on Reddit, Stribling writes, "We literally sat around for two weeks and just brainstormed buzzwords, clauses, paragraph structures and other paper elements just based on what we thought would be funny. That's the grammar. Then SCIgen itself just goes through the grammar and makes random choices to fill stuff in. That's why you see things like 'a testbed of Gameboys' in the evaluation sections sometimes—we just thought it would be hilarious."[11]

The creators intended for SCIgen "to maximize amusement."[12] Laughing at a SCIgen paper expresses opposition to the appropriation of open access ideals for profitability, the gaming of academic metrics at all levels, and reading's disappearance from evaluations of scholarly merit. Despite an apparent lack of seriousness, humor critiques and highlights problems

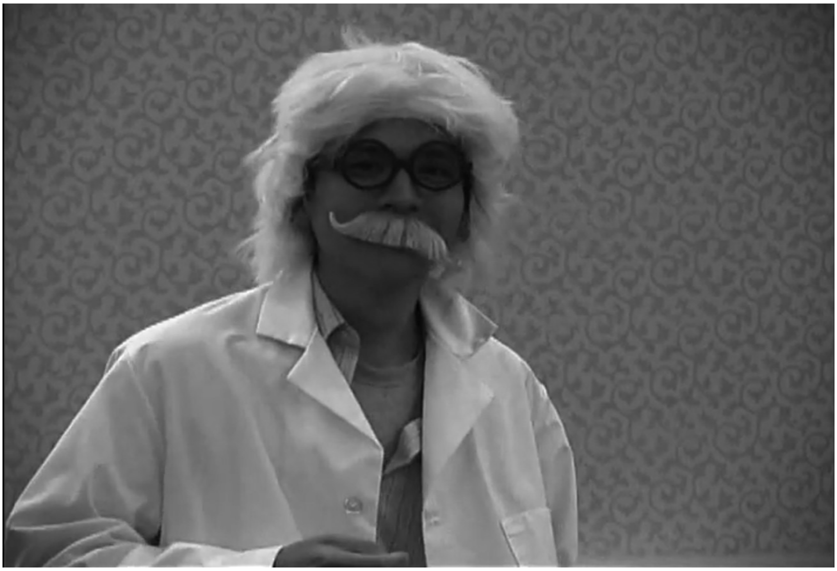


Figure 21.1
Aguayo presenting as Franz T. Shenkrishnan, PhD.

within academia. SCIgen's creators felt "fed up with all of the bogus journals and conferences that spam researchers and charge crazy fees for articles they don't even read."[13]

Against the original intent of its creators, SCIgen also proved popular amongst the journals, conferences, and scholars they intended to mock. Counterfeit journals populate their websites with SCIgen text and scholars submit SCIgen papers to conferences and journals with the hopes of getting published. Labbé discovered SCIgen has been used to generate hundreds of published conference papers (Van Noorden, 2014; Bohannon, 2015) and to game Google Scholar's rankings with self-citing SCIgen papers (Antkare, this volume, chapter 14). After academic publisher Springer's embarrassment that their journals had accepted more than 100 SCIgen papers, the publisher hired Labbé to create SciDetect, a computer program to spot SCIgen content.

The founders of SCIgen, however, thought that SciDetect facilitated the continuation of not reading. For Stribling, SciDetect "seems like just a way for them to avoid having real peer review.…There are better ways to solve the problems exposed by SCIgen than just having a detector."[14] Relying on SciDetect incorporates the technology of critique but not the point of critique: the rise of spammish scholarship and the end of reading in evaluation.

Reading also played a role in one of the SCIgen developer's early projects. As a college student at Harvard, Krohn co-created TheSpark.com, which became SparkNotes. With its cheat sheets summarizing books, SparkNotes promises, "When your books and teachers don't make sense, we do." In their Ask Me Anything session on Reddit, Krohn mentions TheSpark when asked about his favorite programming language: "The original version [of SCIgen] was programmed in…Perl! I ripped the code off from TheSpark.com's high school English paper generator, which was also written in Perl. It's since been modernized. All of the magic is in the grammar rules though." Like TheSpark promised to eliminate the need for reading the original texts, SCIgen promises to eliminate the need for writing original texts. Both projects relish cracking and reverse engineering text.

While both the Postmodernism Generator and SCIgen generate nonsensical articles as a method for critique, the Postmodernism Generator created articles that poked fun at figures of academic authority. By contrast, the primary target of SCIgen's critique is not the academic elite but rather fraudsters. SCIgen mocks spammish conferences and journals, shoddy reviews, easily manipulated rankings, and metrics-driven academic misconduct. The different targets mark a shift in perceived threats to science from postmodern theorists in the 1990s to spammish science in the 2000s.

Conclusion

Changes in humor reflect changes in the ecology of scholarship. Sometimes these jokes embrace vulgarity. The same year that WMSCI accepted the SCIgen paper for their conference, David Mazières and Eddie Kohler wrote a paper titled "Get me off Your Fucking Mailing List," [*sic*] which they submitted to WMSCI. The paper consists almost entirely of the sentence, "Get me off your fucking mailing list," repeated many times, including in two illustrative figures that simply restate, "Get me off your fucking mailing list." The paper was not accepted.

Nearly ten years later, "Get me off Your Fucking Mailing List" [*sic*] fared better. After receiving an email invitation to submit to the *International Journal of Advanced Computer Technology,* engineer Peter Vamplew sent the "article as a reply to the spam email without any other message."[15] To his surprise, Vamplew—who was not even listed as an author on the paper—received notice of the paper's acceptance for publication "with minor changes" along with a boilerplate reviewer report and a request for $150 to be paid to the account of Tej Pal Singh.

In 2013, John Bohannon conducted what he called a "sting operation" by submitting a fatally flawed paper about the "anticancer" properties of a chemical extracted from lichen to *Science*. Bohannon submitted versions of the paper under pseudonyms such as Ocorrafoo Cobange, a biologist at the fictional Wassee Institute of Medicine in Asmara. Of the 304 submissions, 157 of the journals accepted the article, revealing—for Bohannon—the "contours of an emerging Wild West in academic publishing" (2013). In this Wild West, journals take advantage of open-access article processing charges to earn money from scholars overeager, even desperate, to publish. While Bohannon identified "predatory" open-access journals as primarily a problem of the global South (India, in particular), the issue is in fact transnational and widespread. Even journals published by Elsevier, Wolters Kluwer, and Sage accepted the hoax article apparently without carefully reading it.[16]

In his essay "The Death of the Author," Roland Barthes proposed a separation between text and writer. Unlike the author who precedes and nourishes his book, the modern writer or scriptor, Barthes suggested, is "born simultaneously with his text" (1977). The Dada Engine and SCIgen may appear to fit the bill as scriptors in which author and text come into being precisely at the same moment. Yet the paper generators' creators, who wrote the code, the grammar, and chose the language—creating possibilities for "a testbed of Gameboys," continue to precede the text. A native English-speaking person coming across phrases like "a testbed of Gameboys" in a standard academic article would laugh and realize that the article was a joke. Yet hundreds, if not thousands, of SCIgen articles have been published. Rather than the death of the author, what SCIgen announces and pokes fun at is, in fact, the death of the reader. These new players of misconduct detection then shift their critiques away from the authorities of the academy and target the criminal, the counterfeit, and the cut-and-paste scientist.

Yet, as we laugh at their jokes, at biologists with names like Ocorrafoo Cobange, at Tej Pal Singh requesting $150 to publish "Get me off Your Fucking Mailing List," at readers who do not understand the humor of "a testbed of Gameboys," do we realize why we are laughing? Are we laughing at nonnative English speakers or scholars from the global South? Does our laughter at spammish scholarship reaffirm the legitimacy and superiority of powerful publishers who can ask for a $10,000 article processing charge? Laughter judges but offers little analysis, and so we must ask what judgments lie behind our laughter.

Notes

1. It will be interesting to see what cases of misconduct this public model catches compared with the traditional behind-closed-doors model.

2. Hoaxes that target figures of authority continue. For instance, philosophers Philippe Huneman and Anouk Barberousse submitted a hoax article to *Badiou Studies* under the fictitious pseudonym Benedetta Tripodi from the Universitatea Alexandru Ioan Cuza. They explain, "The parody is designed to undermine the foundations on which the ontology of the 'Master' rests, its use to determine how social relations work, how radical politics can be based, and, apart from anything else, is highly amusing." Accessed June 1, 2017. http://retractionwatch.com/2016/04/07/philosophy-journal-spoofed-retracts-hoax-article/.

3. As in our Introduction, the term "spammish" points to the new misconduct's affinities with "spam." Labels such as "fake" and "predatory" fail to capture the complexity of these new forms of scholarly manipulation.

4. Gabriella Coleman has also written about how humor, jokes, and even funny code figure prominently in hacker discourses (2013).

5. The founder was anonymous for the first three years. Anonymity, which makes some uncomfortable or dismissive, can allow for more honest discussion among scientists. Accessed July 5, 2016. http://science.sciencemag.org/content/341/6146/606.full.

6. "PubPeer FAQ." Accessed July 2, 2016. https://pubpeer.com/faq.

7. Watchdogs' anonymity can anger scholars accused of misconduct, who sometimes sue to unmask, and then sue their critics. Unlike traditional carnival, in which the authorities permit mocking as a way to reestablish hierarchy and status quo, scholars who are mocked may fight back. This lack of permissiveness around mocking might result from the fact that the stakes are different. There can be consequences after the carnival. The status quo is not always reestablished. Scientists can lose their jobs. http://retractionwatch.com/2015/04/13/lawsuit-involving-pubpeer-unmasks-commenter-as-pseudonymous-whistleblower-clare-francis/.

8. Science Fraud is archived on the Internet Archive's "Way Back Machine."

9. They note that, "all but two of the top thirty are men, which agrees with the general findings of a 2013 paper suggesting that men are more likely to commit fraud." Accessed July 5, 2016. http://retractionwatch.com/the-retraction-watch-leaderboard.

10. They documented their prank with this video. Accessed October 1, 2016. https://www.youtube.com/watch?v=uT-WOSI2tXg.

11. "At MIT we created SCIgen, which generates gibberish science papers that continue to fool academic conferences. Ask us anything!" Accessed April 8, 2019. https://www.reddit.com/r/IAmA/comments/32l0ym/at_mit_we_created_scigen_which_ generates/.

12. "SCIgen - An Automatic CS Paper Generator," https://pdos.csail.mit.edu/archive/scigen/.

13. These quotations are taken from SCIgen's "Ask Me Anything" subreddit. https://www.reddit.com/r/IAmA/comments/32l0ym/at_mit_we_created_scigen_which_ generates/.

14. Ibid.

15. "Bogus Journal Accepts Profanity Laced Paper," https://scholarlyoa.com/2014/11/20/bogus-journal-accepts-profanity-laced-anti-spam-paper/.

16. For a map of Bohannon's experiment in open access, see http://scim.ag/OA-Sting.

References

Bakhtin, Mikhail. 1984. *Rabelais and His World.* (Helen Iswolsky, trans.) Bloomington: Indiana University Press.

Barthes, Roland. 1977. *Image-Music-Text.* (Stephen Heath, trans.) New York: Farrar, Straus and Giroux.

Biagioli, Mario. 2016. "Watch Out for Cheats in Citation Game." *Nature News* 535, no. 7611 (July 14, 2016): 201.doi:10.1038/535201a.

Bohannon, John. 2013. "Who's Afraid of Peer Review?" *Science* 342(6154):60–65.

Bohannon, John. 2015. "Hoax-Detecting Software Spots Fake Papers." *Science* 348(6230):18–19.

Bulhak, Andrew C. 1996. "On the Simulation of Postmodernism and Mental Debility Using Recursive Transition Networks." Monash University, Department of Computer Science Technical Reports.

Bulhak, Andrew C. "acb." Accessed April 8, 2019. http://dev.null.org/acb.

Bulhak, Andrew C. "Communications from Elsewhere." Accessed April 8, 2019. http://www.elsewhere.org/journal/pomo.

Bulhak, Andrew C. "The Dada Engine." Accessed April 8, 2019. http://dev.null.org/dadaengine/manual-1.0/dada.html#SEC48.

Coleman, Gabriella. 2013. *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton, NJ: Princeton University Press.

Dawkins, Richard. 1998. "Postmodernism Disrobed." *Nature* 394:141–143.

Sokal, Alan. 1996. "A Physicist Experiments with Cultural Studies." *Lingua Franca* May/June.

Stribling, Jeremy, Max Krohn, and Dan Aguayo. "SCIgen—An Automatic CS Paper Generator." Accessed April 8, 2019. https://pdos.csail.mit.edu/archive/scigen.

Van Noorden, Richard. "Publishers Withdraw More Than 120 Gibberish Papers." *Nature News*. Accessed June 7, 2016. doi:10.1038/nature.2014.14763.

Acknowledgments

This book is a direct outcome of the UC Davis Innovating Communication in Scholarship Project and reflects all the conversations, suggestions, and questions from our team: MacKenzie Smith, Jonathan Eisen, Allison Fish, and Alessandro Delfanti. Our colleagues in the Science and Technology Studies Program, the Center for Science and Innovation Studies, and the UC Davis School of Law have had substantial unquantifiable impact on this project, while the support provided by the UC Davis Office for Research's IFHA's Program has been equally generous but clearly quantifiable. Special thanks go to Harris Lewin.

An earlier conference on "Gaming the Game" curated by Kriss Ravetto-Biagioli and Colin Milburn at UC Davis inspired the thinking that went into organizing our 2016 "Gaming Metrics" meeting from which this book has grown. We wish to thank Jeffrey Beall, Carl Bergstrom, John Bohannon, Johan Bollen, Anupam Chander, Martin Kenney, Tim Lenoir, Karen Levy, Dan Morgan, Lior Pachter, Madhavi Sunder, Darren Taichman, and Debora Weber-Wulff for their comments and interventions that made that conference a memorable one.

Geof Bowker's skillful yet unobtrusive guidance and the candid comments by three anonymous reviewers have helped to morph a set of interesting papers into an organized whole, which has then assumed book form thanks to Paul Edwards, Katie Helke, and Justin Kehoe. We hope you are all as proud of the outcome as we are. Lastly, while it is said that information wants to be free, it surely would have missed that goal without a grant from the UC Davis Library's TOME Program to support the OA version of this book. Thank you.

Contributors

Boris Barbour is a research director at École des Neurosciences Paris Île de France and a board member of the nonprofit The PubPeer Foundation.

Mario Biagioli is a Distinguished Professor in the School of Law and Department of Communication at UCLA.

Paul S. Brookes is a professor in the Department of Anesthesiology and Perioperative Medicine and in the Department of Pharmacology and Physiology at the University of Rochester.

Finn Brunton is an assistant professor in the Department of Media, Culture, and Communication at New York University.

Alex Csiszar is an associate professor of the history of science at Harvard University.

Alessandro Delfanti is an assistant professor of culture and new media at the University of Toronto.

Emmanuel Didier is a founding member and permanent researcher at the joint CNRS-UCLA lab Epidapo (Epigenetics, Data, Politics).

Sarah de Rijcke is a professor of science, technology, and innovation studies and deputy director at the Centre for Science and Technology Studies at Leiden University.

Daniele Fanelli is a fellow in quantitative methodology in the Department of Methodology at the London School of Economics and Political Science.

Yves Gingras is a professor and Canada research chair in history and sociology of science, Department of History, at Université du Québec à Montréal.

James Griesemer is a professor and chair of the Department of Philosophy at the University of California, Davis.

Catherine Guaspare is a sociologist at the CNRS (National French Research Center) and an affiliate at the joint UCLA-CNRS lab Epidapo.

Marie-Andrée Jacob is a professor of law at the University of Leeds.

Barbara M. Kehm is a fellow at the Leibniz Center for Science and Society of the Leibniz University Hannover.

Cyril Labbé is Maître de Conférences at the Université Grenoble Alpes.

Jennifer Lin is director of project management at CrossRef.

Alexandra Lippman is a visiting assistant professor at Pitzer College.

Burkhard Morgenstern is a professor of bioinformatics at Universität Göttingen.

Ivan Oransky is a distinguished writer in residence at New York University's Carter Journalism Institute and president of the Association of Health Care Journalists.

Michael Power is a professor of accounting at the London School of Economics.

Sergio Sismondo is a professor in the Department of Philosophy at Queen's University, Canada.

Brandon M. Stell is a co-founder of The PubPeer Foundation and a researcher at CNRS.

Tereza Stöckelová is a researcher at the Institute of Sociology of the Czech Academy of Sciences and assistant professor at the Department of General Anthropology, Charles University.

Elizabeth Wager is a freelance consultant and trainer in medical publishing and peer review and founder of Sideview.

Paul Wouters is a professor of scientometrics and director of the Centre for Science and Technology Studies at Leiden University.

Index

Bold page numbers refer to figures

Abnormal Science, 170
Academia.edu, 4, 262, 265, 267–268
Academy of Science Prize, 159
ACRank, 247
Aggarwal, Bharat, 170–171
Aguayo, Dan, 275–**277**
Allison, David, 141–142
Altmetrics, 38, 50, **52**, 139, 213–224
American Association for the Advancement of Science (AAAS), 48
American Civil Liberties Union, 143
American Heart Association, 169
American Journal of Pure and Applied Mathematics, 244–245
Annapolis Group, 96
Anonymity, 124, 169–172, 280n5
  publishing pressure and, 111
  PubPeer and, 151–153, 159–160
  watchdog groups and, 17, 137–138, 142, 164, 256, 272–274, 280n7
Antkare, Ike, 12, 84, 139, 177–189, 239, 261, 274, 277
Archives des Sciences, 263
Aristotle, 65
Article-Level Metrics (ALM), 213–215
ArXiv, 51, 154–155, 178, 188, 190n16, 240, 261–268
Audit culture/society, 1, 42, 44, 64, 67, 73, 79
Australia, 93, 95, 115
  cash bonuses for publication, 6, 114
Australian, The, 93
Authors, 8, 50, 85–86, 91, 149–150, 172, 181, 206, 259n1
  altmetrics and, 214, 233
  arXiv and, 265–266
  author fees, 13, 47, 204
  cash bonuses to, 6
  citations and, 36, 69, 85, 130, 263
  co-authors, 47, 92, 102–104, 114–118, 116, 159–163, 275
  COPE and, 257
  creating metric-enhancing evidence, 1
  death of the author, 279
  exploitation of, 249
  fake authorship, 2, 14, 84, 92, 139, 142, 145, 179, 185, 188, 239, 274, 278
  fake peer review by, 9, 12, 20n22, 49
  gift authorship, 229
  hoaxes and, 240, 252–255, 273
  industry-sponsored articles by, 125–132
  institutional requirements for, 10, 18n1
  as key opinion leaders, 124–127
  post-publication peer review and, 137, 152–154
  prior evaluation models for, 5
  purchased authorship, 7
  relationship to reviewers, 16
  spammish publishing by, 15
  symbolic capital of, 45

Badiou Studies, 280n2
Bakhtin, Mikhail, 241, 273–274
Baltimore, David, 173
Barberousse, Anouk, 280n2
Barbour, Boris, 137, 201
  See also Vigilant Scientists
Barthes, Roland, 279
Baudrillard, Jean, 275
Baulcombe, David, 159–160
Beall, Jeffrey, 9, 13
  predatory journals list, 19n9, 92, 103, 108n6, 108n8, 109n14, 186, 249n1
Biagioli, Mario, 67, 71, 79, 224, 243, 259
Bik, Elisabeth, 144
Bilder, Geoffrey, 224
BioRxiv, 154–155, 240
Birmingham University, 57–58, 63
Boghossian, Peter, 255, 259n2
Bohannon, John, 252–253
Bonanza, 139, 209, 274
Bourdieu, Pierre, 43, 45
Brand appropriation, 239, 261, 264, 267–268
Brandtstädter, Susanne, 258
Brazil, 11, 173
Brincat, M., 127
British Times Higher Education, 94, 99
Brookes, Paul S., 138, 272
  See also Science Fraud
Brunton, Finn, 105, 124, 239–240, 256
Bulhak, Andrew C., 274–275
  See also Postmodernism Generator
Bureau of Education, 99

Cabanac, Guillaume, 190n18
Canada, 93, 96, 99
Carnivalesque, 17, 137–138, 240–241, 271–275
Cartwright, Hoss, 139, 209–211, 274
Casadevall, Arturo, 145
Catelli, James, 99
Cattell, Dinah, 129–130
Cell magazine, 9, 160, 164n1
Center for Advanced Studies in the Behavioral Studies (CASBS), 31–32, 35
Center for Higher Education Development (CHE), 96
Center for Research Quality and Policy Impact Studies (R-QUEST), 74
Chamayou, Grégoire, 78
Chamberlain, Scott, 220
Charles University, 102
  Ethical Commission, 104
Chen, Shuyang, 185
China, 7, 97, 185, 254, 258
  cash bonuses for publication, 6, 49, 114
Circulation Research, 151
Citation Flow, 247
Citations, 13, 107, 145–146, 155, 179, 223, 239
  altmetrics and, 50, 214–216
  citation rings, 2, 9–12, 16, 91–92, 243–249
  counting process, 262–263
  evaluation metrics based on, 3–4, 48–49, 63–64, 69–73, 267
  fake, 183–185, 188, 190n9, 275
  gaming, 1, 31–39, 44, 81, 85, 129, 139, 213
  infrastructure of, 67, 70, 106
  prediction of, 6
  reasons for, 202–203
  self-, 10–11, 130–131, 189n1
  study of, 46–47
  as symbolic capital, 45
Citefactor, 13
CiteULike, 214
Clarivate Analytics, 179
  Journal Citation Reports, 10–11
  See also Thomson Reuters
Cobange, Ocorrafoo, 279
Cochran, Johnny, 169
Cole, Jonathan and Stephen, 33, 36
Coleman, Gabriella, 280n4
Collaboration Score, 143
Cologne University, 96

Comfort, Alex, 36
Committee on Publication Ethics (COPE), 12, 240, 251, 256–258
Competition, 3, 96–99, 150, 169–170, 262, 266
  for author fees, 13
  for evaluation metrics, 27, 70–71
  for funding, 102, 149
  journal creation and, 47
  in science, 158–159, 202, 254
  for symbolic capital, 43
Computer-generation, 12, 139, 177, 239, 271–272
Contract research organizations (CROs), 123
Copying, 98, 239–240, 254, 256, 258, 261–268
  See also Dafen paintings
Cornell University, 240, 262
Costas, Rodrigo, 113, 115
Crisis in scholarly publishing, 43
Crossref, 139
  Event Data, 222
Csiszar, Alex, 8, 27, 78, 123
Curi, Rui, 173
Current Trends in Technology and Sciences, 208
Curricula vitae (CVs), 79, 86, 117, 216
  fake, 103, 139, 204–205, 208–211, 253
  padding, 3, 9
  as reason for publishing, 10, 15, 92, 127, 132, 246, 248
Czech Grant Agency, 102
Czech Republic, 101–102, 106, 107n3, 109n11
Czech Science Foundation, 107n1

Dada Engine, 178, 274–275, 279
Dafen paintings, 254, 258
Dalkilic, Mehmet M., 188
Daston, Lorraine, 80
Data manipulation, 2, 9, 140, 160–164, 215–220, 219, 231, 272
Dawkins, Richard, 274
Debord, Guy, 178, 275
Delfanti, Alessandro, 240
De Rijcke, Sarah, 71, 74, 91, 208, 248, 254
Designwrite, 127
Didier, Emmanuel, 138
Dill, David, 98
Disney, 204–205, 208–209
Distance University of Hagen, 96
Donmez, Gizem, 172–173

East European Scientific Journal, 244
Economics and Society, 108n5
Editors, 15–16, 20n22, 36, 255, 257
  acknowledgements sections and, 126
  altmetrics and, 214
  editorial boards, 9, 14, 92, 103–104, 106, 201–211, 244
  editorial management systems/workflows, 145, 153
  editorial policy, 150–151
  fake, 108n5, 139, 239, 244, 274
  gaming metrics, 1–2, 9–11, 44
  ghostwriting and, 127
  peer review by, 46, 185
  PubPeer notifying, 160
  retractions by, 49, 112
  selling papers, 19n17
  Sokal affair and, 259nn1–2, 271
Eigenfactor, 216
Einstein, Albert, 139, 181, 274
Eisen, Jonathan, 85, 206–207
Elkana, Yehuda, 32
Elsevier, 47, 179, 279
  See also Scopus
EMBO Journal, 160
Erkkilä, Tero, 97
Erwin, Joseph, 211
Espeland, Wendy, 57
European Commission, 98, 105
European Review of Social Sciences, 108n5
European U-Multirank Consortium, 97–98
European Union, 101, 107n1
Ezrahi, Yaron, 33, 40n16
Ezzat, Sheeran, 273

Fabrication, 229, 255, 257
  of audit trails, 64
  of author information, 12–13
  of data, 111, 116, 152, 159, 162, 164, 171, 231
  defining misconduct, 2–3, 8
Facebook, 204–206, 208, 214, 244, 246
Fake conferences, 275–278
Fake/spammish conferences, 9, 11–15, 178, 272
Falsification, 153
  of data, 51, 229, 231, 257
  defining misconduct, 2–3, 8
  of metrics, 177
  See also Fraud; Plagiarism
Fanelli, Daniele, 92
Fang, Ferric C., 145
Finance (newspaper), 99
Financial Times, 6
Food and Drug Administration (FDA), 129
Foucault, Michel, 62, 65, 178, 275
France, 99, 138, 159–160, 186, 189n2, 273
Francis, Clare, 172
Franssen, Thomas, 74
Fraud, 65, 162, 252, 278
  altmetrics and, 218
  defining misconduct, 2–3
  detection of, 144, 158, 255
  fraudulent articles, 1, 157
  fraudulent image manipulation, 231
  fraudulent journals, 15
  fraudulent metrics, 247
  fraudulent peer review, 49
  gendered, 280n9
  mocking, 241, 271–273
  retractions for, 146, 185
  rewards for, 142, 164
  spam and, 244
  traditional *vs.* post-production, 9
  See also Falsification; Plagiarism; Science Fraud
Freeland, Richard, 13
French National Center for Scientific Research (CNRS), 159, 161–164
Freud, Sigmund, 181

Galison, Peter, 80
Garfield, Eugene, 27, 33, 35–37, 69
Genes & Development, 160
George Washington University, 173
German Excellence Initiative, 97, 107n1
German Research Funding Agency (DFG), 172
German Society for Sociology, 96
German Society of Chemists, 96
German Society of Education, 96
German Society of Historians, 96
Germany, 96–97, 99, 107n1, 139, 172, 204–205, 208–209
Germany in Search of the Super Star, 96
Ghost-managing, 123–132
Gibbs, Philip, 265
  See also viXra
Giles, C. Lee, 188
G-index, 181, 183
Gingras, Yves, 28, 248
Global Impact Factor (GIF), 13
Globalization, 254
Global Journal of Pure and Applied Mathematics, 245
Global North, 15, 91, 239, 248
Global South, 91–92, 251–259, 279
Goal displacement, 27, 33–34, 36–38
Goldacre, Ben, 143
Goodhart's law, 28, 37–39, 78–86, 123, 216
  definition, 1
  as economic concept, 31
  origin of, 18n2
  variations of, 62
Google
Google PageRank, 245–246
Google Scholar, 70, 179, 262, 267, 274, 277
  Ike Antkare and, 139, 177, 181–184
Gorry, Philippe, 129
Goudsmit, Samuel, 36

Great Britain, 97, 99
  See also United Kingdom
Greene, Sarah, 143
  See also Collaboration Score
Griesemer, James, 28–29, 189
Griliches, Zvi, 33
Guardian, 99
Guarente, Leonard, 172
Guaspare, Catherine, 138
Guttenberg, K. T., 205

H2020 Project PRINTEGER, 107n1
Hacking, 2, 8, 28–29, 77–86, 218, 246
Hamburg University, 96
Handelsblatt, 96
Harvard University, 93, 99, 278
Hazelkorn, Ellen, 95, 99
Healy, David, 129–130
Herbert Open Access Journal Biology, 208
Hermès, 186–**187**
Higher Education Council for Education, 221
Higher Education Funding Council for England
  The Metric Tide, 71
H-index (Hirsch index), 70, 80, 85, 118, 141, 216
  Ike Antkare's, 181, 183–184
Hiring, 5, 13, 16, 20n19, 103, 129, 150, 277
Hm-index, 181, 183
Hoaxes, 108, 239, 241, 252–256, 259n2, 271–279, 280n2
Hofstadter, Douglas, 274
Holton, Gerald, 33
Hoskin, Keith, 38
Huang, Tao, 188
Hughes, Raymond, 98
Humor, 12, 17, 137–139, 239–240, 252, 271–279
Huneman, Philippe, 280n2

Image manipulation, 144, 158, 161, **171–173,** 231, 272
Impact Case Study (ICS), 58–65
Impact or perish, 1, 10, 91
  See also Publish or perish (concept)
Indexes (journal), 14, 36, 69, 105–106, 154
  expanded, 5
  gaming, 13
  impact factor and, 4–6
  performance, 179
  quantitative, 5
  study of, 35
  See also Google Scholar; H-index (Hirsch index); Nature Index; Scholarometer; Science Citation Index (SCI); Scopus
Indonesia, 251
InSPIRE, 262–263, 267
Institut de Biologie Moléculaire des Plantes (IBMP), 159
Institute for Scientific Information (ISI), 13, 27, 33, 36–37, 93, 177, 179, 209
  See also ISI-Web of Knowledge (WoK)
Institute of Electrical and Electronic Engineers (IEEE), 12, 177, 185
International Archives of Medicine, 208
International Conference on Advances in Communication and Computing Technologies (ICACACT), 186
International Economics Letters, 108n5
International Journal of Advanced Computer Technology, 278
International Journal of Agriculture Innovations and Research, 208, 211, 274
International Journal of Applied Science and Mathematics, 208
International Journal of Biotechnology, 208
International Journal of Research and Innovations in Earth Science, 208
International Journal of Research in Agricultural Sciences, 208
Ireland, 99

ISI-Web of Knowledge (WoK), 177, 179, 184
Italy, 99

Jacob, Marie-Andrée, 103, 132, 207, 240, 249
James Cook University, 95
Japan, 107n3
  cash bonuses for publication, 49
Jigen, Jiuchii, 170, 172
Johns, Adrian, 256
Jornal Brasileiro de Pneumologia, 11
Journal impact factor (JIF), 6–9, 18n5, 37, 48, 67, 70, 85, 105, 216
Journal of Advances in Biology, 208
Journal of Agriculture and Life Sciences, 210
Journal of Biological Chemistry, 145
Journal of Biomolecular Structural Dynamics (JBSD), 10
Journal of Primatology, 139, 209–211, 274
Journal of Veterinary Advances, 210

Karin, Michael, 172
Keele University, 57–58, 63
Kehm, Barbara M., 91
King Abdulaziz University, 20nn19
KNOWSCIENCE project, 74
Kohler, Eddie, 278
Krishnamoorthy, Mukkai, 188
Krohn, Max, 275–276, 278
  See also SparkNotes/TheSpark.com
Krugman, Paul, 181
Kruskal, William, 33
Kuhn, Thomas, 3
Kumar, Rakesh, 173

Labbé, Cyril, 12, 84, 139, 274, 277
  See also Antkare, Ike; SciDetect
Labbé, Dominique, 190n18
Laborjournal, 160
LabTimes, 160
Lambert Academic Publishing, 104
Lancet, 146
La Repubblica, 99
Larivière, Vincent, 113, 115
Larkin, Marilynn, 130
La Trobe University, 93
Laura and John Arnold Foundation, 143, 150
Lavoie, Allen, 188
Law, John, 101
Lederberg, Joshua, 32
Lee, Sam W., 173
Leiden Manifesto, 71, 80–81
Leiden University, 113
Leipzig University, 96
Le Monde, 160–162, 164n1
Libraries, 8–9, 46, 48, 84–85, 108n6, 248
  journal subscription prices and, 43, 47
Liew, Foo Y., 172
Lin, Jennifer, 139
Lin, Wen-Yuan, 101
Lindsay, James A., 255, 259n2
Lingua Franca, 259n2
LinkedIn, 214
Linux, 268
Lippman, Alexandra, 67, 71, 224, 240, 243, 259
Logic of auditability, 58, 62–65
London School of Economics (LSE), 60–61
Lucas, Robert, 77–78
Lyotard, Jean-François, 275

MacDougall, Curtis, 255
Maclean, 96, 99
Macmillan, 50
Marcus, Adam, 141
  See also Retraction Watch
Margolis, Joel, 36
Marx, Karl, 181
Massachusetts General Hospital, 173
Mathgen, 179–180
Matthew effect, 95
May, Kenneth O., 36
Mazières, David, 278

MD Anderson, 170
Media, 129, 139, 177, 214, 255, 261, 266
  coverage of misconduct, 104, 117, 161
Melbourne Business School
  cash bonuses for publication, 6
Mendeley, 214, 217–218
Merton, Robert, 27–28, 32–38, 45, 48, 68
Metadata, 4–5, 17, 67, 70, 262, 267
Meta-gaming, 61–62, 77–86
Miami University, 98
Misconduct, definition, 2–3, 8
Misconduct paradigm, 3
MIT, 11, 172, 276, 280n11
  Computer Science and Artificial Intelligence Laboratory, 275
  CSAIL, 178
Molecular Biology Journal, 203–208
Morgenstern, Burkhard, 139, 244, 274
  See also Cartwright, Hoss; Uhnemann, Peter
Morphew, Christopher C., 94
Munro, Margaret, 143

National Information Standards Organization (NISO)
  Alternative Metrics Initiative, 222
Nature Index, 50
Nature magazine, 9, 35–36, 141, 144, 149, 151, 160
  impact factors, 13, 48–50, 108n8
  payment for publication in, 6–7
Navarro-Yashin, Yael, 253
Netherlands, 101, 114
New Knowledge Management, 44
New Public Management, 157
New Scientist, 36
New York University, 271
New Zealand, 115
Neylon, Cameron, 224
Northeastern University, 13
Norwegian Research Council, 74
Nouvel Observateur, 99
NPG, 165

Office of Research Integrity (ORI), 17, 137, 170–171, 271, 272
OMICS Publishing Group, 203–204, 206–207, 209–211
Open access publishing, 43, 44, 105, 252, 265, 268, 276
  gold model, 101
  See also Herbert Open Access Journal Biology; Scholarly Open Access
Open Science Foundation, 150
Oransky, Ivan, 117, 137
  See also Retraction Watch
Oxford University, 57

Pachter, Lior, 20n19
PAK International Journal of Veterinary Sciences, 210–211
Pakistan
  cash bonuses for publication, 49
Peer review, 81, 107, 229, 232, 233, 277
  citations and, 48, 69
  for conferences, 16, 44, 145, 185, 259n2
  credibility of, 1, 15, 71–73, 185
  critiques of, 46
  duped reviewers, 255
  by editors, 46, 185
  fake, 9, 12, 20n22, 49, 91
  fast-track, 51
  hoaxes and, 253, 259n2
  informed, 68, 72–73
  for journals, 103, 230–231, 261–262, 265–266
  lack of, 106, 188, 267, 268
  limitations of, 154
  misrepresentations of, 14
  post-publication, 70, 137, 152–154, 169–175
  publishing pressure and, 186
  qualitative, 47
  study of, 46
  testing, 178
Photoshop, 82, 144, 158
Physical Review, 36

Plagiarism, 11, 205, 229
  defining misconduct, 2–3, 8
  plagiarism detection software, 144, 189n4, 265
  See also Falsification; Fraud
Plant Cell, 161
Plato, 65
PLoS (Public Library of Science), 13, 85, 154
  PLoS ONE (journal), 213–215
Pluckrose, Helen, 255, 259n2
PNAS, 160, 170
Postmedia, 143
Postmodernism Generator, 271–272, 274–275, 278
Power, Michael, 28, 247
Predatory journals, 5, 82, 190nn16–17, 243, 246–248, 254
  Beall's list of, 19n9, 92, 103, 108n6, 108n8, 109n14, 186, 249n1
  impact factors at, 13, 85
  increase of, 44
  international publishing and, 101–107, 279
  mocking of, 241, 262, 272
  open-access publishing and, 47, **52**
  pay-to-play business model, 9
  terminology, 14–15, 20n21, 280n3
  watchdogs targeting, 252
Preprints, 51, 154, 240, 261–262, 265–268
Price, Derek J. de Solla, 36
Promotion (career), 5, 37, 47, 62, 142, 146, 151
  impact factors in, 35, 103–104, 177
Pseudonyms, 17, 138, 169–175, 189n1, 241, 253, 271–276, 279, 280n2
Publishers, 8, 50, 103–104, 154, 177, 185
  altmetrics and, 213–214
  author fees and, 13, 279
  criticisms of, 143
  ethics of, 79
  fake, 203, 208–211
  profitability of, 232
  publication management systems at, 20
  relationship to predatory journals, 14–15, 105–107, 108n6, 108n8, 186, 239
  relationship to tenure system, 49
  retractions and, 16, 145
  SCIgen content and, 188, 278
Publish or perish (concept), 1, 3, 44, 145–146
  See also Impact or perish
Publish or Perish (tool), 181
PubMed, 144, 170
  PubMed Commons, 174
PubPeer, 17, 149–150, 154–155, 159–162, 174, 231
  anonymity on, 142–143, 151–153, 159–160, 273
  discussions on, 144, 172
  ignoring evaluation metrics, 137–138
PuLP (Public Library of Philosophy), 84–86

Qualitative assessment, 5, 7, 58, 63, 68–69, 111, 137
  qualitative peer review, 47
Quantitative assessment, 5, 43, 47–49, 63–64, 69, 78–80
  frameworks for, 221
  limits of, 71–73, 201–202
  new methodologies of, 102
  of retractions, 137
Queensland University Business School
  cash bonuses for publication, 6

Rankings, 44, 47, 50, 63, 73, 262–263, 277–278
  of global universities, 5, 7, 9–10, 13–14, 27, 51, 69, 91, 93–99, 177, 189n2
  of journals, 11, 13, 35–36, 107n3, 114–115, 231
  of retractions, 146, 273
  of scholars, 139, 181, **183**, 275
  of UK universities, 57, 65n1
  of US universities, 1, 57, 94
  of websites, 184, 245–246

Raymond, Eric, 17
Reddit, 214, 217, 276
  Ask Me Anything, 276, 278
Reproducibility, 17, 145, 223, 273
Research Excellence Framework (REF), 57–58, 61–62, 65, 221
ResearchGate, 4, 161, 218, 267
Research Integrity and Peer Review, 85
Retractions, 12, 16, 169, 171, 185, 231
  fake peer review and, 49
  ghostwriting and, 128
  rates of, 44, 112–115, 174, 229
  risk of, 117–118
  watchdogs for, 137–138, 141–146, 160–164, 272–273
  See also Retraction Watch
Retraction Watch, 12, 17, 137–138, 141–146, 160–**162**, 272–273
Riksbanken Jubileumsfond, 74
RIV-points (Information Register of R&D results), 102
ROpenSci, 220
Rushforth, Alexander D., 71
Russia, 99, 104, 108n5

SAGE, 20n22, 279
Salami-slicing, 92, 111, 118–119, 140, 229
Salmi, Jamil, 95, 99
Saroyan, Alenoush, 99
Sauder, Michael, 57
Saudi Arabia, 93
Schneider, Leonid, 160
Scholarly Open Access, 17, 19n9
Scholarometer, 139, 179, 181, **183**
ScholarOne, 20n22
SciDetect, 12, 188, 277
Science Citation Analysis System (SCAS), 37
Science Citation Index (SCI), 27, 33, 47, 69, 123
  See also Web of Science (WoS)
Science Fraud, 138, 143, 272–273, 280n8
Science Immunology, 48
Science magazine, 6–7, 9, 48, 93, 160, 279
Science Robotics, 48
SCIgen, 11–12, 178
Scopus, 47, 70, 103–104, 107, 108n5, 109n14, 177, 179, 184–185
Shanghai Jiao Tong University
  Academic Ranking of World Universities (ARWU), 93–95, 97, 99
Sideview, 139
Siegel, James, 251
Simpson, O. J., 169
Singh, Tej Pal, 278–279
Sismondo, Sergio, 44, 92, 256
Slutsky, Robert, 1
Smith, George and Richard, 273
  See also PubPeer
Smith, Normand, 173
Snark, 138, 169–175
Social media, 51
  part of altmetrics, 139, 214
  research sharing through, 17, 174, 202, 262–263, 267
Social Science Research Council, 32
Social Science Research Network (SSRN), 262, 265, 268
Social Text, 256, 259n2, 271
Sokal, Alan, 255–256, 259n2, 271, 274
South Korea, 49
  cash bonuses for publication, 49
Spain, 184
Spammish journals, 137, 139, 239–240, 272, 274, 277–279
  terminology, 14–15, 243, 280n3
SparkNotes/TheSpark.com, 278
Spiegel, 99
Springer, 12, 50, 177–178, 188, 190n17, 277
  See also SciDetect
SSME, 185
Stack Exchange, 214
Stanford University, 31
Stapel, Diederik, 144
Steen, R. Grant, 145
Stell, Brandon M., 137, 201, 273
  See also PubPeer; Vigilant Scientists

Stöckelová, Tereza, 91, 107n1, 208, 248, 254
Strathern, Marilyn, 37–38
Stribling, Jeremy, 275–276
Students, 4, 7, 33–34, 37, 95, 172, 177, 278
  graduate students, 11, 35, 73, 170, 257, 275–276
  as resource, **46, 52**
  student performance, 8, 14
  university rankings and, 1, 95–96
Sunday Times, 99
Swanson, Christopher, 94
Sweden, 74
Swiss Federal Institute of Technology (ETHZ), 159, 161–164, 165n2
Swiss National Science Foundation, 163
Switzerland, 159, 161–164
Symbolic capital, 45–**46,** 68, 256

Technische Universität München-Institute for Advanced Study, 107n1
Thackray, Arnold, 32
Thomson Reuters, 10–11, 13, 19–20, 179, 209
  See also Clarivate Analytics
Toth, Cory, 142–143
Tripodi, Benedetta, 280n2
Turing, Alan, 181
Turkey, 114
Turkish Republic of Northern Cyprus (TRNC), 253
Twitter, 214, 248
  part of altmetrics, 50, 218
  research sharing through, 214, 217
  spam accounts, 246–247

Uhnemann, Peter, 139, 204–209
UK Institute of Cancer Research, 65n1
Ukkonen, Esko, 201
United Kingdom, 109n11, 114
  research evaluation in, 57–65
  See also Great Britain
United States, 17, 137, 170, 244, 271
  academic production in, 113–115
  federal definition of misconduct, 2
  organizational sociology in, 34
  patents in, 107n3
  pharmaceutical industry in, 126–127
  publishing gatekeepers in, 105
  science in, 31–32
  universities in, 93–95, 97
  university rankings in, 1, 57, 94, 96, 99
Universal Impact Factor (UIF), 13
University of Calgary, 142
University of California, Davis, 206
University of California, San Diego, 1
University of Cambridge, 93, 103
  Energy Policy Research Group, 109n11
University of Colorado, Denver, 108n6
University of Glasgow, 172
University of Illinois, 275
University of Loughborough, 109n11
University of Massachusetts at Amherst, 275
University of Montréal, 113
University of New South Wales, 93
University of Rochester, 272
University of South Carolina, 161
University of Strasbourg, 103, 159–160, 163–164
University of Trieste, 12
US Congress, 32
US National Science Foundation (NSF), 163
  Science Indicators (SI-72), 32–33
US News and World Report, 1, 13, 20n19, 94, 96, 99

Vamplew, Peter, 278
Vance, Vicki, 161–**162**
Van Noorden, Richard, 144
Vigilant Scientists, 152
ViXra, 240, 261–268
Voinnet affair, 138, 157–164, 273

Wager, Elizabeth, 140
Wakefield, Andrew, 146
Walsh, Lynda, 255

Warry Report, 58–59, 63
Wass, Joe, 224
Web of Science (WoS), 11, 47, 70, 103
  See also Science Citation Index (SCI)
WikiLeaks, 138
Wikipedia, 214, 218, 268
Williams, Kyle, 188
Wolters Kluwer, 280
Wong, Winnie, 254
World Multiconference on Systemics, Cybernetics, and Information (WMSCI), 276, 278
World War II, 262
Wouters, Paul, 28, 58, 77, 217
Wulfenia, 263

Xia, Jingfeng, 253
Xiong, Jiping, 188

Yale University, 14

ZEIT, 99
Zotero, 214
Zuckerman, Harriet, 32